43786

From:	Bugaisky, Gabriele
Sent:	Monday, June 04, 2001 10:35 AM
To:	STIC-Biotech/ChemLib
Subject:	09/596784

These are two microbial proteins that are co-expressed from an operon: please search SEQ ID NO:2 and 4, and interference files of both

thanks, gabi

Gabriele E. Bugaisky
- au 1653
- cm1-10d09
- 308-4201

1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:16:39 ; Search time 89.87 Seconds
(without alignments)
2397.109 Million cell updates/sec

Title: US-09-596-784-2
Perfect score: 9448
Sequence: 1 MELKSLGTEHKAAVHTAAHN.........NPQVASALTDLKKEGLEMKS 1838

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   SPTREMBL_15:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	9448	100.0	1838	2	O54581	O54581 erwinia amy
2	5529.5	58.5	1829	2	Q9KH44	Q9KH44 erwinia her
3	1824	19.3	1795	2	Q66101	Q66101 pseudomonas
4	283	3.0	57	2	Q9LAW1	Q9LAW1 erwinia amy
5	275	2.9	2514	2	Q9JY30	Q9JY30 neisseria m
6	255.5	2.7	3455	2	Q9P9U6	Q9P9U6 xylella fas
7	247	2.6	3442	2	Q9PBE8	Q9PBE8 xylella fas
8	245	2.6	2703	2	Q9K0T0	Q9K0T0 neisseria m
9	244.5	2.6	3259	4	Q14789	Q14789 homo sapien
10	240.5	2.5	5327	5	Q76891	Q76891 drosophila
11	240.5	2.5	5476	5	Q9NJ17	Q9NJ17 drosophila
12	240.5	2.5	5533	5	Q9U6C3	Q9U6C3 drosophila
13	240.5	2.5	5554	5	Q9NHN1	Q9NHN1 drosophila
14	237.5	2.5	4152	2	Q9ZHL3	Q9ZHL3 haemophilus
15	236.5	2.5	3012	5	Q97205	Q97205 leishmania
16	229.5	2.4	2273	2	Q31152	Q31152 neisseria m
17	229	2.4	3381	5	Q9KX33	Q9KX33 streptococc
18	229	2.4	2540	4	Q9NZQ2	Q9NZQ2 homo sapien
19	222	2.3	4823	13	Q93321	Q93321 fugu rubrip
20	220.5	2.3	2253	13	P70012	P70012 xenopus lae
21	220	2.3	494	2	Q52389	Q52389 pseudomonas
22	220	2.3	4498	13	O93291	O93291 fugu rubrip
23	219.5	2.3	2614	5	O97054	O97054 dictyosteli
24	218.5	2.3	4957	4	O14687	O14687 homo sapien
25	218.5	2.3	5262	4	O14686	O14686 homo sapien
26	217	2.3	2712	10	Q9SB74	Q9SB74 arabidopsis
27	216.5	2.3	3634	2	Q9JP78	Q9JP78 bordetella
28	214	2.3	3282	2	Q9PEY9	Q9PEY9 xylella fas
29	212.5	2.2	1557	2	Q9RNI2	Q9RNI2 haemophilus
30	212.5	2.2	2478	2	Q9LCH2	Q9LCH2 staphylococ
31	212	2.2	2324	11	Q9WUJ3	Q9WUJ3 rattus norv
32	211	2.2	1794	9	Q9T1A7	Q9T1A7 bacteriopha
33	211	2.2	2647	5	Q9U4X0	Q9U4X0 plasmodium
34	210.5	2.2	2478	5	Q9RL69	Q9RL69 staphylococ
35	210	2.2	2059	5	Q9PD50	Q9PD50 xylella fas
36	210	2.2	2151	5	Q9NG79	Q9NG79 trichomonas
37	209.5	2.2	1510	5	O61802	O61802 caenorhabdi
38	209.5	2.2	2178	2	Q9KWR3	Q9KWR3 streptococc
39	209.5	2.2	4919	2	Q9ZHL0	Q9ZHL0 haemophilus
40	207.5	2.2	3590	2	Q45365	Q45365 bordetella
41	206.5	2.2	2406	5	Q9VXM5	Q9VXM5 drosophila
42	206	2.2	3257	5	Q9V736	Q9V736 drosophila
43	205.5	2.2	4545	2	Q9X4W2	Q9X4W2 vibrio chol
44	204.5	2.2	2110	5	Q9VRA6	Q9VRA6 drosophila
45	204.5	2.2	4558	2	Q9KS12	Q9KS12 vibrio chol

ALIGNMENTS

RESULT 1
O54581
ID   O54581   PRELIMINARY;   PRT;   1838 AA.
AC   O54581;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   DSPE.
GN   DSPE OR DSPA.
OS   Erwinia amylovora.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Erwinia.
OX   NCBI_TaxID=552;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EA321;
RX   MEDLINE=98115919; PubMed=9448330;
RA   Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
RA   Conlin A.K., Collmer A., Beer S.V.;
RT   "Homology and functional similarity of an hrp-linked pathogenicity
RT   locus, dspEF, of Erwinia amylovora and the avirulence locus avrE of
RT   Pseudomonas syringae pathovar tomato.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EA321;
RA   Bogdanove A.J., Kim J.F., Wei Z.-M., Kolchinsky P., Beer S.V.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CFBP1430;
RX   MEDLINE=98086111; PubMed=9426142;
RA   Gaudriault S., Malandrin L., Paulin J.P., Barny M.A.;
RT   "DspA, an essential pathogenicity factor of Erwinia amylovora showing
RT   homology with AvrE of Pseudomonas syringae, is secreted via the Hrp
RT   secretion pathway in a DspB-dependent way.";
RL   Mol. Microbiol. 26:1057-1069(1997).
RN   [4]
RP   SEQUENCE OF 1-238 FROM N.A.
RC   STRAIN=EA321, ATCC 49947;
RX   MEDLINE=98422475; PubMed=9748455;
RA   Kim J.F., Beer S.V.;

```
RT   "HrpW of Erwinia amylovora, a new harpin that contains a domain
RT   homologous to pectate lyases of a distinct class.";
RL   J. Bacteriol. 180:5203-5210(1998).
DR   EMBL; U97504; AAC04850.1; -.
DR   EMBL; Y13831; CAA74156.1; -.
DR   EMBL; U94513; AAC62315.1; -.
SQ   SEQUENCE   1838 AA;  198243 MW;  7D595F78130E8FD9 CRC64;

Query Match             100.0%;  Score 9448;  DB 2;  Length 1838;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1838;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1 MELKSLGTEHRAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MELKSLGTEHRAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP 60

Qy   61 STAADGISAAHQQKKSFSLRGCLGTKKFSRSRAPQGQPGTTHSKGATLRDLLARDDGETQH 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 STAADGISAAHQQKKSFSLRGCLGTKKFSRSRAPQGQPGTTHSKGATLRDLLARDDGETQH 120

Qy  121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180

Qy  181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240

Qy  241 LHQQRLARERENPPQQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 LHQQRLARERENPPQQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300

Qy  301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360

Qy  361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420

Qy  421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480

Qy  481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540

Qy  541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL 600

Qy  601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660

Qy  661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720

Qy  721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780

Qy  781 AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT 840

Qy  841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG 900

Qy  901 GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV 960

Qy  961 SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020

Qy 1021 HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080

Qy 1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140

Qy 1141 KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADAKLVSDHKPDADQIKQLRQQF 1200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADAKLVSDHKPDADQIKQLRQQF 1200

Qy 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260

Qy 1261 SAELAKKLKNTLLSLDSGESMSFSRSRYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 SAELAKKLKNTLLSLDSGESMSFSRSRYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320

Qy 1321 LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380

Qy 1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440

Qy 1441 FSVDFTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 FSVDFTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500

Qy 1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560

Qy 1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620

Qy 1621 DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680

Qy 1681 AALPASSAKRLGEMMNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 AALPASSAKRLGEMMNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740

Qy 1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800

Qy 1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
         ||||||||||||||||||||||||||||||||||||||
Db 1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838


RESULT    2
Q9KH44
ID   Q9KH44      PRELIMINARY;      PRT;   1829 AA.
AC   Q9KH44;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   DSPE.
```

```
GN   DSPE.
OS   Erwinia herbicola pv. gypsophilae.
OC   Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
OC   Erwinia.
OX   NCBI_TaxID=48984;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mor H., Manulis S., Zuck M., Nizan R., Coplin D.L., Barash I.;
RT   "Genetic Organization of the hrp Gene Cluster and dspEF Operon in
RT   Erwinia herbicola pv. gypsophilae.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF271717; AAF76343.1; -.
SQ   SEQUENCE 1829 AA;  201195 MW;  CD81C57122EE2470 CRC64;

Query Match             58.5%;  Score 5529.5;  DB 2;  Length 1829;
Best Local Similarity   59.2%;  Pred. No. 0;
Matches 1103;  Conservative 268;  Mismatches 430;  Indels 63;  Gaps 14;
```

```
Qy      3  LKSLG--TEHKAAVHTAAHNPVGHGVALQQGSSSSSSPQNAAASLAAEGKNRGKMPRIHQP    60
           :|:|:|  |||||||:|||||: :|||||||||| ||||||||||||||| ||
Db      1   MKLIGHLTEQKTVVQNVERTVYDAGTTLQQGSSCSAFQAAGLLTSEGKNRGKMPNIRQY     60

Qy     61  STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHS-----KGATLRDLLAR-    113
           :: |:||  |||||||||:| |  |::|::  :|  |  :||       |||||
Db     61  FKGENKKDAARQQKKSFSLKNLFEFKKSAKPPEQTSNATATAREMGIRRRPFLGELFAQP   120

Qy    114  DDGETQHEAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQ   173
           :| :::  :|  |||||||| |:|| :: :|  |  |  ::|:    :  | |:|||||
Db    121  EDDDSTEPLSSPLARRLTRSVGIRRDSMKELQGRPVLSADQDVD---HHVKQQLHNFRE    176

Qy    174  MRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQED   233
           :||::|  |  :| |:|  :|  |   :|    ||           ||          ::
Db    177  VRQSILNGIGQ-SSSSAGKTPLTEPAEAPAS-------AVPSTLS--------EITEEN   219

Qy    234  DDSEFQQLHQQRLARERENPQPPPKLGVATPISARFQPKLTAVAESVLEGTDT-------   286
           |||||:|:|||||||:|:||:||||    ||| | |||| |  :|:|| ::
Db    220  DDSEFEQMHQQRLARERNNPLPPPN---GPVARQFQPTLSPVAQSLPKAETSQKLLLLQ   275

Qy    287  TQSPLKPQQSMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHA   346
           ||||||        :|:| :||:|:|:|:|||| :  ||:|||||:|||||| ::||: :|
Db    276  TQSPLK------RVPVSPIGITIDNGKLQLGLGGSNLPAINTLLEQTLGKDCQRYLTHSA   327

Qy    347  SSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISL   406
           ::| :||||||||:|||:|:|:| ||||||||: :||: :|| || ||  :: ::|||
Db    328  SADSRQHLLLDNKGRLFDIISHEGGYSVLHNSQSSSLRTQLTQNEDTPVRLENDNGKITI   387

Qy    407  GSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAA-RPQGESIRLHDDKIHILHPELGVWQ   465
           | | : :|: :|:||:: : :L|:GIWQH |||  | :|| :|||||||||:| :||:|:
Db    388  SLGAEGQNRLNLNQPGDVQHALISGIWQHSTAAGMQSENEYVRLNDDKIHILNHVIGVWK   447

Qy    466  SADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEEKLVDKIKSYSVDQRGQVAILTDT   525
           :: :  :SQLSRQ|DGKLYA::D RTL:NLSDN::SE::|VDKIKS:SV :RGQVAILTDT
Db    448  ASGNTPYSQLSRQGDGKLYAVQDGRTLHNLSDNHVSEKFVDKIKSFSVSERGQVAILTDT   507

Qy    526  PGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGK   585
            | |: :|MPS  | |  : :SLH:AD: : L:::|:||:| | :|||:|||::||||||
Db    508  DSPHHLCLMPSTKAPAGQRIQFSLHLADTMLMLQRGESHLDAQSIAISNERLYAADSEGK   567

Qy    586  LFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHAC   645
           |:|  : : :DG EL ::  :Q:|L::HFGHDH:I GFF D HGQLNAL:K:NFRQQHAC
Db    568  LYSGVLSQIRDG-ELPMEQILQKVLNQHFGHDHRIEGFFTDHHGQLNALVKDNFRQQHAC   626

Qy    646  PLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTK   705
           PLGNDHQFH GWNL:D:L I::QLG: :   ||:||::|: |GSL:LQ:GK::YFDQLTK
Db    627  PLGNDHQFHAGWNLSDTLFINSQLGMDNVDTAPHKILNLGREGSLTLQDGKVYYFDQLTK   686

Qy    706  GWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPG   765
           GW: AES C:QLKKGLDG AY:LK:GEVKRL:IN:STSSI :G :N:F:LPHVRNKPEPG
Db    687  GWSSAESGCQQLKKGLDGGAYILKEGEVKRLDINHSTSSISQGKDNFFALPHVRNKPEPG   746

Qy    766  DALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGEL   825
           :||QGL:K |KAQA:AVIGVN:YLAL::KGDI:S:QI:PGTQQL RP::TLSREG:SG L
Db    747  SALQGLEKADKAQAIAVIGVNHYLALSDKGDIKSYQIRPGTQQLARPTRTLSREGLSGTL   806

Qy    826  KDIHVDHKQNLYALTHEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKP   885
           KDIH:DH:QNLYA::H:GE:FHQ RE WQNG  SS W KLA:P ::S L  L:M  :H:P
Db    807  KDIHIDHEQNLYAVNHDGEIFHQSREQWQNGKPSSGWQKLAMPHNDSDLDRLEMDGDHQP   866

Qy    886  IATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKL   945
           :AT::DGS::QL: : WHA  :  |  L VG S SQ::F RL::GVKG:::PG:G:TVK
Db    867  VATLVDGSRYQLRGKAWHAQTSVEPTPLEAGLSNSQQIFVRLNRGVKGRLLIPGTGVTVKA  926

Qy    946  SAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYE  1005
           SAQ : :T  E  K:SSKF::R:RA: AFNPTM TPRP:KNAAY  QH WQGR:GL  :YE
Db    927  SAQFMAQTYRENWKIRSKFADRVRAHIFNPTMTTPRPLKNAAYNLQHRWQGRKGLSAVYE   986

Qy   1006  MQGALIKQLDAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTV  1065
           MQGALIKQL::NVR:S IQ DL SKL::LDLG: : :LLNDMKRFR:ELEQSATRS :
Db    987  MQGALIKQLESQNVRNSTIQTDLLSKLDALDLGQGTSLNDMKRFREELEQSATRSAIM   1046

Qy   1066  LGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQS  1125
           LGQHQGVL:SNG:IN:SFKPS  KA::QSFNVNRSG:DLSK:L:QAV:A  PS :SKL::
Db   1047  LGQHQGVLNSNGQINNSFKPSRSKAAIQSFNVNRSGRDLSKALEQAVLAAAPSPQSKLRT  1106

Qy   1126  MLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADAKAKLVSD  1185
           :L HFVSAG::MS::KG:IP:GRQRDPND:T:LTKSRLILDTVT:GELH:LAD::A:LVS:
Db   1107  LLSHFVSAGLNMSYKKGDIPMGRQRDPNDQTSLTKSRLILDTVTLGELHQLADKAALVSG   1166

Qy   1186  HKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEH  1245
           HKP: :Q::QLRQQFD:LREK:Y :NPVK::TDMG:THN:ALEANYDAVKA INAF KEH
Db   1167  HKPYPNQVQQLRQQFDALREKHYGANPVKQFTDMGLTHNTALEANYDAVKACINAFSKEH   1226

Qy   1246  HGVNLTTRTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPV  1305
           HGVNLTTRTVLE:QG  :L KKLK:TLLSLD GES:SFSR:YGG :STVF PT::KKVPV
Db   1227  HGVNLTTRTVLETQGNPALEKKLKDTLLSLDVGESLSFSRAYGGLSTAFAPTI-KKVPV   1285

Qy   1306  PVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGVSGNIMVATGHDVMPYMTGKKTSAGN   1365
           P::PGAGITLDRAYNLSF RT:G L:V FGRDG:: :::V:TG:D:MPYMTGKKT:A:N
Db   1286  PIVPGAGITLDRAYNLSFGRTTGALEVGFSRDGGATGTVSVSTGRDLMPYMTGKKTIADN  1345

Qy   1366  ASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELL  1425
           ASDWLS KHKISPDLRIGA::S :L: TLQNSL FKLTEDELP F::GL::GTL PVEL:
Db   1346  ASDWLSKKHKISPDLRIGATLSSNLKRTLQNSLSFKLTEDELPSFLKGLIQGTLNPVELM  1405

Qy   1426  QKGIEHQMKQGSKLTFSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGS  1485
           QKGIEHQM:QGS:L F:VD SA LDL AGIN:   | ::   :  R S GL:ASANL :
Db   1406  QKGIEHQMRQGSQLVFNVDVSAALDLYAGINITDCVHESDTLIFRASTGLRASANLVTAS  1465

Qy   1486  RERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEG----------KP  1534
           R R: : G: :::SASNNRPTFLN  :AGAN ::  G|A:   ::              :|
Db   1466  RARSKQRGELSNTVSASDNRPTFLNSAALGANTTLNAGVANDFIREASINDDGTQKPEKK  1525

Qy   1535  VGIFPAFTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKM  1594
           VG FPAF S NV::L:A:DNRT:QSIS:E:K:AEPVT NDI EL:STL KHFKDSA:TK
Db   1526  VGTFPAFVSPNVTVSLAMDNRTTQSISVEMKQAEPVTLNDIGELVSTLAKHFKDSASTKF  1585

Qy   1595  LAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSHMELGSA  1654
           L: L::  :  PA:QL :L :HFS  :V:GD:RYEAVR LK:L ::QQA :S:M:LG:A
Db   1586  LSDLNKQQNPDPAKQLSDLLRHFSGSNVMGDDRYEAVRRLKQLGQRQQAYESNMMKLGTA  1645

Qy   1655  SHSTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFS  1714
           :H :Y:NLSR::::GIV:LL :HF AALP::SA  L :MM::DP LK :IK LQ: PFS
Db   1646  RHITQYSNLSRLDENGIVDLLRHHFTAALPSTSATSLSKMMDRDPLLKSLIKLQTKPFS  1705

Qy   1715  SASVSMELKDGLREQTEKAILDGKVGREEVGLFQDRNNLRVKSVSVSQSVSKSEGFNTP  1774
           :ASVSMELKD LR::TEKAIL:GK:GR:E:G:LFQDRNNLR::S:SVSQ:V:KSEGFN P
Db   1706  NASVSMELKDDLRDKTEKAILEGKIGRDELGVLFQDRNNLRISSISVSQNVTKSEGFNMP  1765

Qy   1775  ALLLGTSNSAAMSMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASALTDLKKEGL  1834
           AL:L ::NSA :SM R IGTINF YGQDQ TPRRFTL   IA:AN :VASAL::LKKEG
Db   1766  ALILSASNSAGLSMGRIIGTINFIYGQDQETPRRFTLAGEIAKANAEVASALSELKKEGF  1825

Qy   1835  EMKS  1838
           ||||
Db   1826  EMKS  1829
```

us-09-596-784-2.rspt

```
RESULT   3
066101
ID   066101          PRELIMINARY;          PRT;     1795 AA.
AC   066101;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   AVIRULENCE PROTEIN (AVRE).
GN   AVRE.
OS   Pseudomonas syringae (pv. tomato).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=98115919; PubMed=9448330;
RA   Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
RA   Conlin A.K., Collmer A., Beer S.V.;
RT   "Homology and functional similarity of an hrp-linked pathogenicity
RT   locus, dspEF, of Erwinia amylovora and the avirulence locus avrE of
RT   Pseudomonas syringae pathovar tomato.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=98422476; PubMed=9748456;
RA   Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA   Collmer A.;
RT   "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
RT   to harpins and pectate lyases and can elicit the plant hypersensitive
RT   response and bind to pectate.";
RL   J. Bacteriol. 180:5211-5217(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=20243785; PubMed=10781092;
RA   Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
RA   Petnicki-Ocwieja T., van Dijk K., Collmer A.;
RT   "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT   mosaic structure composed of a cluster of type III secretion genes
RT   bounded by exchangeable effector and conserved effector loci that
RT   contribute to parasitic fitness and pathogenicity in plants.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
DR   EMBL; U97505; AAC06134.1; -.
DR   EMBL; AF232006; AAF71499.1; -.
SQ   SEQUENCE  1795 AA;  195351 MW;  260F74534DE08D5F CRC64;

Query Match            19.3%;  Score 1824;  DB 2;  Length 1795;
Best Local Similarity  29.1%;  Pred. No. 5.4e-98;
Matches  554;  Conservative 358;  Mismatches 722;  Indels 272;  Gaps 71;
```

```
Qy      34 SSSPQNAAASLAAEGKNRGKMPRIHQPSTAADGISAAHQQKKSFSLRGCLGTKKFSRSAP 93
                                                   :: |||||   ||||| ||
Db      61 SKAPQQKAAT-PPTAKN------VKTPPPASNVATPRNKARES--------GFSNSSP 103

Qy      94 Q---------GQPGTTHSKGATLRDLLARDDGETOHEAAAPDAARLTRSSGGVKRRNMD 142
                         ||     || :|:   |||
Db     104 QNTHRAPKWILRNHPNQASSSGAQTHEI---------HPEAAP-----------RKNLR 142

Qy     143 DMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIP 202
                                                              :| | || :|
Db     143 VRFDLPQ------------------------------------DRLERSPSYLD 160

Qy     203 GSHHEIKEEPVGSTSKATTAHADRVEIAQEDDSEFOQLHQQRLARERENPQOP---PK 258
            ||||      ||   :|     |  |||                           ||
Db     161 SDNPMTDEEAVANATRQF-------RSPDSHLQGSDGTRISMLATDPDQPSSSSGSK 209

Qy     259 LGVAT-PISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGSGAGVTPL-AVTLD-KGK 315
            :| ||  || || ::  |        :  ||       |     :| || ||| | |||
Db     259 NLTTPFIILADK--ATGLWPTAGATGNRNYILNAERCEGGVTLYLISEGA-GNVSGGF-- 315
```

```
Db     210 IGDSDGPIPPR-----------EPMLWRSNGGRFELKDEKLVRNS----EPQGSIQLDAKGK 256
               | | | ||              :|| |  | ||:||||||||      ||:||||| |||
Qy     316 LQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVL 375

Db     257 PDFSTFNTPGLAPLLDSLLATPKQTYLAHQSKDGVHGHQLLQANGHFLHLAQDDSSLAVI 316
            ||:|::|  ||| ||| ||| |: | ||| :  | :||:|| :| |||:| ||||| ||
Qy     376 HNSHPG-EIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQ 434

Db     317 RSSNEALLIEGKKPPA------VKMEREDGNIHI-DTASGRKTQ-ELPGKAHIAHITNV-- 367
            :|::  |||  |            |::||| |:|  : | |||  ||| || | || |
Qy     435 HPAGAARPQGESIRLHDKIHILHPELGVWQSAD--KDT-HSQLSRQADGKLYALKDNRT 491

Db     368 ---LLSHDGERMRVHEDRLYQFDPISTRWKIPEGLEDTAFNSLSTGGNGSVYAKSDDAV 423
                || ||||| |:|:||||| | |:| |||||||||| |||| ||| |: |:|:| |
Qy     492 LQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHF 551

Db     424 V-DLSSPFMPHVEVEDLQSFSVAPDNRAALLS----GKTTQAILLT-DMSP----VIGGLT 474
            : | ||| |  ||:||  |: |    | |||     ||| | :|| |||      ||| |
Qy     552 ADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDNELKMKAMPQHALD 611

Db     475 PKKTKGLELDGGKAQAAAVGLSGDKLFIADTQGRLYSAD-RSAFEGDDPKLKLMPEQANF 533
            |  ||||| | |  | | || | ||| |||||||| ||  | |||| | ||||||| |:
Qy     612 E-----HFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLG-NDHQFHPGWNLTDALVID 666

Db     534 QLEGVPLGGHNRVTGFINGDDGGVHALIKNRQGETHSHALDEQSSKLQSGWNLTNALVLN 593
            |||| || |||| |  | |  || |::||||  | | ||| || |||| |||| |||: |
Qy     667 NQLGLHHTNPEP--HEILDMGHLGSLALQEGKLHYFDQLTKGWTGAE-SDCKQLKKGLD 722

Db     594 NNRGL--TMPPPTAADRLNLDRAGLVGLSEGRIQRWDATPECWKDAGIKDIDRLQRGAD 651
            |||      | || ||:|:|||| || ||| ||: || | | || ||| |:||||| |
Qy     723 GAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKA-QAMA 781

Db     652 SNAYVLKGGKLHALKIAAEHPNMAFDRNTALAQTARSTKVEMGKEIEGL--DDRVIKAFA 709
            | ||||:|| | | | |||| ||| || ||| |  ||:|| |||| || |   | |:|||
Qy     782 VIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALTH 841

Db     710 MVSNKRFVALDD--------QNKLTAHSKDHKPVTLDIPGLEGDIKSLSLDEKHNLHALTS 762
            :| ||:|| ||         |||| | ||| |||| |:|| || |||||||:| |||| |
Qy     842 EGEVFHQPREAWQN--GAE-SSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQH-- 895

Db     763 TGGLYCLPKEAWQSTKLGDQLRARWTPVALPGGQ-PVKALFTNDDNVLSAQIEDAEGKGL 821
            ||| | |||||||| |||||| |||| || ||   |||| |||||||||||:| ||||||
Qy     896 -QLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPSGLTVKLSAQTGGMTG 954

Db     822 MQLKAGQWQRF--EQRP--VEENGLNDVHSRITGSNKTWRIPKTGLTLRMDVNTFGRSG 876
            ||||| ||:| |||    |||| |||| ||:||||| ||||| || || |||||||| |
Qy     955 AEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQL 1014

Db     877 VEKSKKAST-SEFIRANIYKNTAEFTPRWMKNVGDHIQHRYQRLGLKEVYETESMLFKQL 935
            |:| |||   |||:|| | | |  | |||  |||| |:|||||||| |: |::|||||||
Qy    1015 DAHNVRHNAPQP-------DLQSKLETLD--LGEHGAELLNDMKRFRDELEQSATRSVTVLG 1067

Db     936 EL--IHESGGRPPARGQDLKARITALEAKLGPQGATLVKELETLRDELENHSYTALMSIG 993
            ||   |||| |  |||| ||| ||| ||||| | | ||||| |||| |:| || |:|| |
Qy    1068 -------------QHGVLKSNGEINSEFKPSP---GKALVQ---SFNVNRSGQDLSKSLQQ 1110

Db     994 QSYCKAKNLKQQDGILNQHGEL---AKPSVRMQFGFKKLADLGTKLNFKSSGHDLVKELQD 1050
            ||  ||:| |  || :|::| |     |:|| | | ||| || | | :|||| |||:| ||
Qy    1111 AVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLLDTVTI 1170

Db    1051 ALTQVAPSAENPTKKLLGTLKHQGLKLSHQKADIPLGQRRDASEDHGLSKARLALDLVTL 1110
            | ||:|||  ||||| | ||| |||||||||||||||:|||| | | |||||||||||||
Qy    1171 GELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEAN 1230

Db    1111 KSLGALLDQV---EQLPPQSDIEPLQKKLATLRDVTYGENPVKVVTDMGFTDNKALESG 1166
            | ||| ||      |||   |:|  |||| || |:|| ||||| |||||| |||||| |
Qy    1231 YDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLLSLDSG-ESMSFSRSYGG 1289

Db    1167 YESVKTFLKSFKKADHAVSVNMRAATGSKDQAELAGKFKSMLKQLEHGDDEVGLQRSYGV 1226
            | |||||| || | ||||| | | | || ||||||| |:|||:|| | |: |:|||||
Qy    1290 GVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLN---VSFGRDGGVSGNIMV 1346

Db    1227 NLTTPFIILADK--ATGLWPTAGATGNRNYILNAERCEGGVTLYLISEGA-GNVSGGF-- 1281
            ||||||:||||    ||||| || |||||||| |||| ||||| ||||| | ||||||
```

```
Qy   1347  ATGHDVMPYM-----TGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFK  1401
                :||  :  |||  :  |:: |: |:  |: |
Db   1282  GAGKDYWPGFFDANNPARSVDVGN------NRTLTPNFRLGVDVTATVAASQRAGVVFN  1334

Qy   1402  LTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLITFSVDTSANLDLRAGINLNED-  1460
           || ::::|| |||| | | | :||::: ||||||||: |||||||| ||||||||||:
Db   1335  VPDEDIDAFVDDLFEGQLNPLQVLKKAVDHESYEARRFNFDLTAGGTADIRAGINLTEDR  1394

Qy   1461  --GSKPN--GVTARVSAGLSA--SANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGA  1514
             ||||        : |||    ||| |||||||  ||||||||||||||||| |||||
Db   1395  DPNADPNSDSFSAVVRGGFAANITVNLMTYTDYSLTQKNDKTELKEGGKNRPRFLNNVTA  1454

Qy   1515  GANLTAALGVAHSSTHEGKPV---GIFPA--FTSTNVSAAL--ALDNRTSQSISLELKRA  1567
                  ||| ||  ||  ||     || || | |  |  ||  || |: ||| |||:||
Db   1455  GGQLRAQIGGSHTAP-TGTPASAPGPTPASQTAANNLGGALNFSVENRTVKRIKFRYNVA  1513

Qy   1568  EPVTSNDISELTSTDLGKHFKDSATTKMLAALKELDDA-----KPAE---QLHILQQHFS  1618
           :|: ::|:|::  |||  | ||: |: |:||| || || :      |||  ||:||||:|
Db   1514  KPITTEGLSKLSKGLGEAFLDNTTKAKLAELADPLNARYTGKKPDEVIQAQLDGLEELFA  1573

Qy   1619  ---AKDVVGDERYEAVRNLKKLVIRQQA-ADSHSMELGSASHSTTYNNLSRINNDGIV-  1672
              ||| |: ||:::|||||| ||:  |||  : :::|||: :  |||||||  :|:|:
Db   1574  DIPPPKD--NDKQYKALRDLKRAAVEHRASANKHSV-MDNARFETSKTNLSGLSSESILT  1630

Qy   1673  ELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEK  1732
               ||  ||  ||   | |: ||  : |  || |:  |:||:| ||||||:| |:|:
Db   1631  KIMSSVRDASAP-GNATRVAEFMRQDPKLRAMLKEMEGSIGTLARVRLEPKDSLVDKIDE  1689

Qy   1733  AILDGKVGREEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNI  1792
               :| |: ||||   |||| :  |:|:|: |:|||| :| :| |||  : |
Db   1690  GSLNGTWTQSDLSSMLEDRNEMRIKRLVVFHTATQAENFTSPTPLVSYNSGANVSVTKTL  1749

Qy   1793  GTINFKYGQDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS  1838
           | |||  |  ||||  ||: |  ::|: |: ||  ||||||| |:|
Db   1750  GRINFVGADQDKPIGYTFDGELSRPSASLKEAAGDLKKEGFELKS  1795
```

```
RESULT  4
ID   Q9LAW1              PRELIMINARY;      PRT;     57 AA.
AC   Q9LAW1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HRP-SECRETED PATHOGENICITY/AVIRULENCE PROTEIN DSPE (FRAGMENT).
GN   DSPE.
OS   Erwinia amylovora.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Erwinia.
OX   NCBI_TaxID=552;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EA246;
RA   Kim J.F., Laby R.J., Beer S.V.;
RT   "Comparison of the hrpN-flanking regions of two Erwinia amylovora
RT   strains with different host specificity.";
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF083620; AAF63403.1; -.
FT   NON_TER        57     57
SQ   SEQUENCE   57 AA;   5788 MW;  AFF0CA36311E4BE1 CRC64;

Query Match             3.0%;  Score 283;  DB 2;  Length 57;
Best Local Similarity  98.2%;  Pred. No. 3.7e-10;
Matches   56;  Conservative   1;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1  MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRI  57
          ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     1  MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGEMPRI  57
```

RESULT 5

```
Q9JY30
ID   Q9JY30              PRELIMINARY;      PRT;   2514 AA.
AC   Q9JY30;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HEMAGGLUTININ/HEMOLYSIN-RELATED PROTEIN.
GN   NMB1768.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002526; AAF42109.1; -.
DR   TIGR; NMB1768; -.
SQ   SEQUENCE   2514 AA;   265615 MW;  95643A671B3BC268 CRC64;

Query Match             2.9%;  Score 275;  DB 2;  Length 2514;
Best Local Similarity  18.88%;  Pred. No. 6e-07;
Matches  389;  Conservative  287;  Mismatches  803;  Indels  586;  Gaps  91;

Qy     7  GTEH-KAAVHTAAHNPV------------------GHGVALQOGSSS----------SSPQNAAAS  43
                ||||||| |:                     || |::|             ::|  ||
Db    37  GSAHVKSVPFGTTHAPVCRSNIFSFSLLGFSLCLAVGTANIAFADGIIADKAAPKTQQAT  96

Qy    44  LAAEGKNRGKMPRIHQPSTAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPG-----  98
                |:||||  |     |  | : :||        |||        ::  ||||  |
Db    97  ILQTGNG--IPQVNIQTPTSAGVSVNQYAQFDVGNRGAILNN--SRSNTQTLGGWIQG  151

Qy    99  -----------------------TTHSK-------------GATLRDLLARDDGETQHEAAAPDAARLTR  132
                                  ||||:|              :|::|
Db   152  NPWLARGEARVVVNQINSSHSSQMNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATL  211

Qy   133  SGGVKRRNMDDMAGRPMVKG-----GSGED------KVPTQOKRHQLNNFGQMRQTMLS  180
                        ::            :||:|       ||   |
Db   212  TTGQPQYQAGDLSGFKIRQGNVVIAGHGLDARDTDFTRILSYHSKIDAPVWGQDVRVVAG  271

Qy   181  KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDSEFQQ  240
                 |||                  :||   |   |: |
Db   272  QNDVVATGNA----HSP------------------------ILNNAAANTSNNTANNGTHIPLFAIDTG-----  312

Qy   241  LHQORLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS  300
                                      ||:   :|:|
Db   301  -------------------------KLG-----GMYANKITLISTA------EQAGIRNQGQLFAS  342

Qy   301  GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHAS--SDGSQHLLLDN  358
             |                 :|             |:|::  :| |::
Db   343  SGN------VAIDANGRLVNSGTMAAAN-----AKDTDNTAEHKVNIRSQG------VEN  385

Qy   359  KG----------HLFDIKSTATSYS---VLHNSHPGEIKGKLAQAGTGSVSVDGKSGKI  404
                          :|:|| |::::|    :|:|||  || |:| |  |::  ||| |:|
Db   386  SGTAVSQQCTQIHSQSIQNTGTLLSSGEILIHNS-----GSLKNETSGTI---EAARL  435

Qy   405  SLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVW  464
                                  || : |     | :| |   : |: |:|
Db   436  AIDTDT-LNNQGKLSQTG---------------------------------SQKLHI--------------------  458

Qy   465  QSADKDTHSQLSRQADGKLYALKDNRTLQNLSD------------NKSSEKLVDKIKSYSVDQRGQ  518
             ||:                                          :| :|      |:  ||:
```

```
Db   459  --------------------DAQGKM---DNRGRMGLQDTAPTASNGSSNQTGN---SYNASFHSS  498

Qy   519  VAILTDPGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLV  578
Db   499  TTTPTTATGTGTATVSISNIFAPT-------FAD---GTIRTHGALDNSGSIIANGQTD  547

Qy   579  VADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN-  637
Db   548  VSAQQG-LNNAG---QIDIHQLNAKG---SAFDNHNG--TIISDAVHIQAGSLNNQNGNI  598

Qy   638  NFRQQHACPLGNDHQFHPGWNLTDALVID-NQLGLHHTNPEPHEILDMGHLGSLALQEGK  696
Db   599  TTRQQ-------------------LEIETDQLDNAHGKLLSAEIADLAVSGSLNNQNGE  638

Qy   697  LHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLP  756
Db   639  IATNQQL-----------------IIHDGQQSTAVIDNTNGTIQSGRDVAIQAK  675

Qy   757  HVRNKPEPGDALQGLNK-------DDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLE  810
Db   676  SLSN----NGTLAADNKLDIALQDDFYYERNIVAGNE-LSLSTRGSLKNSHTLQAGKRIR  730

Qy   811  RPAQTLSREG----ISGELKDIHVDHKQNLYALTHEGEVFHQPREAWQNGAESSSWHKLA  866
Db   731  IKANNLDNAAQGNIQSGGTTDIGTQHN-----LTNRGLIDGQ------  767

Qy   867  LPQSESKLKSLLDMSH--------EHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTS  918
Db   768  ----QTKIQAQCMNNIGTGRIYGDNIATAATRLDNQDENGTGA-AIAARENLNLGIG--  819

Qy   919  GSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMS  978
Db   820  QLNNRENSLIYSGNDMAVGGALDTNGQATGKAQR------  853

Qy   979  TPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQL----------------DAHNVRHNAP  1024
Db   854  ----IHNAGATIEAAGKMRLGVEKLHNTNEHLKTQLVETGREHIVDYEAFGRHELLREGT  909

Qy  1025  QPDL-------QSKLETLDLGEH----------GAELLNDMK  1049
Db   910  QHELGWSVYNDESDHLRTPDGAAHENWHKYDYEKVTQKTQVTQTAPAKIISGNDLTIDGK  969

Qy  1050  R-FRDELEQSATRSVTVLGQHQG-----------VLKSNGEINSEFKPS-PGKALVQSFN  1096
Db   970  EVFNTDSQIIAGGNLIVQTEKDGLHNEQTFGEKKVFSENGKLHSYWREKHKGR---DSTG  1026

Qy  1097  VNRSGQDLSKSLQQAVAHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKT  1156
Db  1027  HSEQNYTLPEEITRNISLGSFAYESHRKALSHHAPSQGTELPQSNG---ISLPYTISNSFT  1083

Qy  1157  ALTKSRL-ILDTVTIGELHELADKAK-----LVSDHKPDA--DQIKQLRQQFDTLREKR  1207
Db  1084  PLPSSSLYIINPVNKGYLVETDPRFANYRQWLGSDYMLDSLKLDPNNLHKRLGDGYYEQR  1143

Qy  1208  YESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTRTVLESQGSAELAKK  1267
Db  1144  LINEQIAELT-GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIAL----SAEQVAQ  1197

Qy  1268  LKNTLLSLDSGESMSFSRSYGGGVSTVFPTLSKKVPVPVIPGAGITLDRAYNLSFSRTS  1327
Db  1198  LTSDIVWLVQKE----VKLPDGTQTVLVPQVYVRVKNGDIDGKGAL------LSGSNTQ  1247

Qy  1328  GGLNVSFGRDGGVSGNIMVATGHDVMPYWTGK---KTSA-------GNASDWLSAKHKIS  1377
Db  1248  INVSGSLKNSGTIAGRNALIINTDLDNIGGRIHAQKSAVTATQDINNIGGMLSARQTLL  1307

Qy  1378  PDLRIGAAV-SGTLQGTLQNSLKFKLTEDELPG-FIHGLTHGTLT----PAELLQKGIE  1430
Db  1308  --LNAGNNINSQSTTASSQNTQGSSTYLDRMAGIYITGKEKGVLAAQAGKDINIIAGQIS  1365

Qy  1431  HQMKQGSKLTFSVDTSANLDL-----RAGINLNEDGSKPNGVTARVSAGLSASA------  1479
Db  1366  NQSEQG-QTRLQAGRDINLDTVQTSKHQATHFDADNHVIRGSTNEVGSSIQTKGDVTLLS  1424
```

```
Qy  1480  --NLAAGSRERSTTSQQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGI  1537
Db  1425  GNNLNAKAAEVSSANGTL--AVSAKND----IN ISAGINT---HVDDASKHTGRSGG-  1473

Qy  1538  FPAFTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTL------GKHFK  1587
Db  1474  ---GNKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANILGSSNVISDNGTQIAAGNHVR  1530

Qy  1588  DSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSH  1647
Db  1531  IGTT-------QTQSQSETYHQTQKSGLMSAGIG---FTI---GSKTN  1565

Qy  1648  SMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMM-----NNDPALK  1702
Db  1566  TQENQSQSNEHTGSTVGSLKGDTTI-VAGKHYE------QIGSTVSSPEGNNTIYAQ  1615

Qy  1703  DIIKQLQSTPFSSASVSMELKDGLREQFEKAILDGKVGREEVGV-----LFQDRRNLRV  1756
Db  1616  SIDIQAAHNKLNSNTTQTYEQKGLTVAFSSPVTD--LAQQAIAVAQSSKQVGQSKND-RV  1672

Qy  1757  KSVSVSQS--VSKSEGFNTPALLGTSNSAAMSMERNIGTINFKYGQDQNTPRRFT-LEG  1813
Db  1673  NAMAAANAGWQAYQTGKSAQNLANGTTNAKQVS-------ISITYGEQQN--RQTTQVQA  1723

Qy  1814  GIAQANPQVASALTDLKKEGLEMKS  1838
Db  1724  NQAQASQIQAGGKTTLIATGAAEQS  1748
```

```
RESULT  6
Q9P9U6

ID   Q9P9U6        PRELIMINARY;      PRT;   3455 AA.
AC   Q9P9U6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HEMAGGLUTININ-LIKE SECRETED PROTEIN.
GN   XF2775.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C. de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawaski H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
```

RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-157(2000).
DR   EMBL; AE004082; AAF85560.1; -.
DR   INTERPRO; IPR000267; -.
DR   INTERPRO; IPR001424; -.
DR   PROSITE; PS00087; SOD_CU_ZN_1; UNKNOWN_1.
DR   PROSITE; PS00144; ASN_GLN_ASE_1; UNKNOWN_1.
SQ   SEQUENCE   3455 AA; 360947 MW; 4CD692CB8752FDAA CRC64;

Query Match   2.7%; Score 255.5; DB 2; Length 3455;
Best Local Similarity 18.6%; Pred. No. 1.4e-05;
Matches 412; Conservative 285; Mismatches 825; Indels 695; Gaps 101;

```
Qy    3 LKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQPST   62
          ::|||      |      ||       ::|        |       ||::||
Db 1374 IDNLGTGRLYGDHIAL-----HAQTLTNRDETSDGHTHAATIAAORLDIGADTLRNTAN 1428

Qy   63 A--ADGISAAHQQKKSFSLRGCL--GTKKF-SRSAPQGQPGTTHSKGATLRDLLARD  114
         |     ||     |  |       ||  ||     || :||
Db 1429 ----GATLDNALHATGTATLLDNRSATIDITGTLNITTTLNNI--RD 1481

Qy  115 DGETQHEAAAPDA--ARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNF  171
         |  :                            :        :
Db 1482 N--VHIAHAPDVVTEARMEQPHW--RKNQPN------GGSGNFRFTSNYDAHDI--- 1525

Qy  172 GQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSK----------A  219
                 ::|     |          :|
Db 1526 -------YYLNPADIIKDD------PYITPDGGQIHRAIVRLTPQTSAYFYARGGLYA 1570

Qy  220 TTAHADRVEIAQEDDDSEFQQLHQQRLARERENPQOPPKLGVATPISARFQPKLTAVAES  279
           |: |||                  |       |        |    :
Db 1571 SQAERRRMDLTARFGDSVVLYYTDR-----QDKQPNPDHVAAA------ATNHS 1613

Qy  280 VLEGTDTTQSPLKPQSMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQ  339
                                    |    |:|
Db 1614 AFIGLDTPQQONERFQT---------VPITYAPGDDRLTYDSNYG----- 1655

Qy  340 HYLAHHASSDSGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDG  399
         |  |            |     ||       |   :
Db 1656 RLVTWHDYTD-PDHTLID---------MHRG-PNDVR------DN 1683

Qy  400 KSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLH--------D  451
         |  |                   :    |  ::|:
Db 1684 EKYR---DATRRTTQQDILNPDAGAPALIQTG------GAMMIQTDTLRNHYADLLAGGD 1733

Qy  452 DKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSY  511
          | :    |                      :|:
Db 1734 QTIVGLPHPTKEKSDDEHKY--------KRVLLIDNRALQ-LS------RTDTFHNI 1776

Qy  512 SVDQRGQVA------------ILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHG  561
         :  ||:
Db 1777 STTYRGKVSEPWSNESRTTPTTQIGGRITSGGHQHIAAQTLNNVTDSTHAPEPIQHLTYN 1836

Qy  562 KSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQIS  621
         |      :|:   : |                  |
Db 1837 PS--TQTLSVVNGVITVTDTSPSLHTVSLADNGPSAGQELTYIPDQSTTTP------ 1885

Qy  622 GFHDDHGQLNALVK----------NN----FRQQHACPLGNDHQFHPGWN  658
         |                               :
Db 1886 --------NAPIRDPAAPPAVTVPTGPLTLPNNSLFTLHPDAATLITTDPRFTLGRP 1935

Qy  659 LTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLT--KGWTGAESDCKQ  716
         | :  |  :               |  |  :
Db 1936 YTFSADTQLHALGDHDT--LHKRLGDGYYEQRLIRE----QLAQLTGRRRLDGYTDDDQQ 1988

Qy  717 LKKGLDGAAYLLKDGEVK------RLNINQSTSSI-----KHGTENVFSLPHVRN  760
         |                  |                 : |
Db 1989 YRALLDAGVTVAKQHLRPGIALSADQLAQLTSDIVWLVQQDVQLPDGTTTRALVPRLYL 2048

Qy  761 KPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTOQLERPAQTLSREG  820
         |         |  |                 :|
Db 2049 RPRTGD----LTPDGALLAAASTTINAH-TFTNTGTIDARHL------IDINAHTMDQQG 2097


Qy  821 ----------ISGELKDIHVDHKQNLYALTH-EGEVFHQP--REAWQNGAESSSWHKL-  865
                          |:  : |              : |
Db 2098 GRLTADAIHIHTGDFTTLGGQFKARGYLKVHAQGNFLASSTLRDATTQGTRHHSVTELD 2157

Qy  866 ----------ALPQSESKLKSLDMSHEHKPIATFEDGSQHOLKAGGWHAYAAPERGPLAVGT-  917
                  :  |  |    |                    ||         :  ||  |||
Db 2158 QQAGFTVTGPGAYLGL-STDQAMTHNGVAINNTGTD------GYTSLNA-TGPLHLGTL 2208

Qy  918 ----------SGSQTVFNRLMQGVKGKVIPGSGLTVKLSA-  947
                    |               || |       ||
Db 2209 TTHRSDTTQWDPRNSRHSRIDTEYGTSITG-NGDIQLNSGQDINLRAATLHSTQGTITAL 2267

Qy  948 QTGGMT-----------------GAEGRKVSSKFSERIRAYAFNPTIMSTPRPIKNAAY  988
                                |  :|:         |           :
Db 2268 ATGNVTIHGDTIQYTSQDSHTKRSGLLNSRTTTHADQQQTQAIGSTLSADKVFVKGNN 2327

Qy  989 ATQHGWQGREGLKPLYEMQGALIKQLDAHNVR--HNAPQPDLQSKLETLDLGEHGAEL 1044
         |         |      |      |  :|           |    :
Db 2328 ITVTGSHVVSDAGTYMQAEHDLTLQAATHTTQSTYSHHTKQRGL------IRNGGASL 2379

Qy 1045 -LNDMKRFRDELEQSATRSVTVLGQHQG-----------VLKSNGEINSEFKPSP 1087
                |      :          :                   :|
Db 2380 TLGNQSQRTDSTTATTTGSLIGATNGNVTLLAGGHYQQIGSDVLSPHGDIDIHAKKVD 2439

Qy 1088 GKALVQSFNVNRSGQDL-----------SKSLQQAVHATPPSAESKLQSMLGH 1129
              |                        :|    | |       :|
Db 2440 ---IIQAHHTSQTTQHTATRQSGLTVALSTPLIAGAQTAQQMQHAAAARSGDPRLQALAGL 2496

Qy 1130 FVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPD 1189
         |               |          :           :
Db 2497 TTALGA------KNTIDAVRQ-DPRALGGLNAS-------LTVG- 2526

Qy 1190 ADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVN 1249
                     |    |:                     :     |
Db 2527 -RSTHDSTTTTTTSTAAGSNVTAGGNVH--ISATGDGTASTLTIQGSDVRG- 2574

Qy 1250 LTTRIVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYG-------GGVSTVFVPTLS 1300
         |   :|                  |     :        |
Db 2575 -DTMTYLKADGDIALL-AAQNTV---TNQRDNRGRSAGVGVAVNLGSGGTSAGLTAHAS 2628

Qy 1301 KKVPVPVIPGAGITLDRAYNLSFSRTSGG-LNVSFGRD--------GGVSGNI 1344
         |                    :   |
Db 2629 TST----GSGHSTDLTW-SNSHVGGGNLLAIDAGGDLLMKGAIGTAKHVIADIAGNL 2680

Qy 1345 MVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLFKLTE 1404
                                     :
Db 2681 RIESLQDTHQYRSTDRSLGGS---------LTAGAGFSGGSANLNHQTIRSDYASV 2726

Qy 1405 DELPGF-------IHGLTH--GTLTP-AELLQKGIEHOMKQGSKLTFSVDTSAN-  1448
         |
Db 2727 TEQSLFTGDGYQLTVGGQTHLIGGAITSNSTAIHNGL-NSLDTGTILQNIENHANYT 2785

Qy 1449 ---LDLRAGINLNE----------------DGSKPNGVTARVS------ 1472
             | :
Db 2786 ATQVNLGGGYSRNGGTVGTDQQGHAATATQVPGTTLPSHNGLSASPPSAMTARDSSHSTT 2845

Qy 1473 -AGLSASANLAAGSRERSTTSGQFGSTTSASNNRP-----  1506
                                      :|
Db 2846 YSGISQGALTIRDDTAQHALTGHTAAETIATLNRDLLTDTATSNALTPIFDEQRINAGFD 2905

Qy 1507 --------TFLNGVGAGANL-------TAALGVAHSSTHEGKPVGIFPAFTSTNVSAALA 1551
                  :                   ||
Db 2906 IVSGLQRETGTFINNRAAEADLKTRQATAADHAAHDPSN---GFNDQQRQTLRDQAIA 2960

Qy 1552 LDNR-------TSQISLELKRAEPVTSNDISELTSTLGKHF------KDSATT 1592
         |                 :           |     |
Db 2961 LTNEAHALKDAWGPGGTYRQITTALAAG--ASGNVSAASSDLAKHMIVNYVQQGATAI 3017

Qy 1593 KMLAALKELDDAKPAE-QLHIL--------------KDV 1622
          | |  |                      ||
Db 3018 GHWVATGQLTEGSPLHAALHALLACAGAAASQQHCSSGAQGAAASSVLTGLFSDPRPEDT 3077
```

```
Qy     3  LKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQPST    62
          :::| || |||  |  |: |  ||: :|    :    |||   | :| : | :| |
Db  1374  IDNLGTGRLYGDHIAL-----HAQTLNRDETSDGHTHAATIAARQRLDIGADTLRNTAN  1428

Qy    63  A---ADGISAAHQQKKSFSLRGCL--GTKKF-SRSAPGQPGTTHSKGATLRDLLARD    114
          |  ||  :||   ::  ::  :|          ||        |::    || ::||
Db  1429  AMILSDGDAAI-----GATLDNALHATGTATLLDNRSATIDITGTLNITTTLNNI--RD  1481

Qy   115  DGETQHEAAAPDA---ARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNF   171
          ||  ||  | |     |  |    :  |    |   |     | | | ||| : | :|
Db  1482  N---VHIAHAPDVVTEARMEQPHW--RKNQPN------GGSGNFRFTSNYDAHDI----  1525

Qy   172  GQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSK-------A   219
                :: :|    : | |    :: | |  :  |  | :  |   :
Db  1526  ------YYLNPADIIKDD-----PYITPDGQQIHRAIVRLTPQTSAYFYARGGLYA  1570

Qy   220  TTAHADRVEIAQEDDDSEFQQLHQQRLARERENPQQPPKLGVATPISARFQPKLTAVAES   279
          :: :  : :::    :  ::       |::          :|
Db  1571  SQAERRRMDLTARTGDSVVLYYTDR-----QDKQPNPDHVAAA-----ATNHS  1613

Qy   280  VLEGTDTTQSPLKPQSMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQ   339
          ::  ||| | |:  |:  |  | :|   :      |        :     : :| :
Db  1614  AFIGLDTPQQNERFQT-----VPITYAPGDDRLTYDSNYG------TCTDDCV  1655

Qy   340  HYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDG   399
          :     :  |  |  ||          ::   : :  | |  :
Db  1656  RLVTWHDYTD-PDHTLID------MHRG-PNDVR---------DN  1683

Qy   400  KSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLH--------D   451
          :  :     :::   :  |  | :|  :      :  ||||  |||: |:|
Db  1684  EKYR---DATRTTQQDILNPDAGAPALIQTG------GAMMIQTDTLRNHYADLLAGGD  1733

Qy   452  DKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSY   511
          :   :|| |: :   :|  :|  :    |  | :|:|| |  |:          :
Db  1734  QTIVGLPPHPTKEKSDDEHKY-----KRVLLIDNRALQ-LS----RTDFFHNI  1776

Qy   512  SVDQRGQVA---------ILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHG   561
          ||  ||:|          || || |
Db  1777  STTYRGKVSEPWSNESRTTPTTQIGGRITSGGHQHIAAQTLNNVTDSTHAPEPIQHLTYN  1836

Qy   562  KSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQIS   621
          :: :|   |:  ||  | | :||| :|: |:| ||||: ||||||   |||||:|||:|
Db  1837  PS--TQTLSVVNGVITVTDTSPSLHTVSLADNGPSAGQELTYIPDQSITTP--------  1885

Qy   622  GFFHDDHGQLNALVK---------NN----FRQQHACPLGNDHQFHPGWN   658
           |                      ||    ||
Db  1886  ---------NAPIRDPAAPPAVTVTPTGPLTLPNNSLETLHPDAATLITTDPRFTLGRP  1935

Qy   659  LTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLT--KGWTGAESDCKQ   716
          :| |: :|| || |||     :  |  | ::   :|::|   :|:|  :
Db  1936  YTSADTQLHALGDHDT---LHKRLGDGYYEQRLIRE---QLAQLTGRRRLDGYTDDDQQ  1988

Qy   717  LKKGLDGAAYLLKDGEVK-----RLNINQSTSSI-------KHGTENVFSLPHVRN   760
          :: |:|  | : |: ::      ||:   :::                :|  |
Db  1989  YRALLDAGVTVAKQHQLRPGIALSADQLAQLTSDIVWLVQQDVQLPDGTTTRALVPRLYL  2048

Qy   761  KPEPGDALQGLNKDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREG   820
          :|  ||  | :  ::  :|   | | :         | ||  ||   |  |    :
Db  2049  RPRTGD----LTPDGALLAAASTTINAH-TFTNTGTIDARHL------IDINAHTMDQQQ  2097

Qy   821  ------ISGELKDIHVDHKQNLYALTH-EGEVFHQP--REAWQNGAESSSWHKL-   865
                |:   :|||:||:||:| || :  ||| | |  |:  |||||||: |: |
Db  2098  GRLTADAIHIHTTGDFTTLGGQFKARGYLKVHAQGNFLASSTLRDATTQGTRHHSVTELD  2157

Qy   866  ------ALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGT-   917
                | ||| ||:| :| : ||:| |||||||| |||||||| |:|| ||| |: |:
Db  2158  QQAGFTVTGPGAVLGL-STDQAMTHNGVAINNTGTD------GYTSLNA-TGPLHLGTL  2208

Qy   918  ------SGSQTVFNRLMQGVKGKVIPGSGLTVKLSA----   947
                |||| || |: |  |:  |||| ||:::
Db  2209  TTHRSDTQWDPRNSRHSRIDTEYGTSITG-NGDIQLNSGQDINLRAATLHSTQGTITAL  2267
```

```
Qy  1623  VGDERVEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGI-----VELLHK  1677
          : || |  :|||:   :| |    ||  | |||||::|:|| ::  :      |||||
Db  3078  AQDR--EAKRNLITSIVTGIASIGHT-DPATATHAA-----IAAVDNNWLAAKQYVQMLNE  3130

Qy  1678  HFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDG  1737
          :|:||   :  ::| |                 |        || :    |  || : ||
Db  3131  EFEAA---TEKEKGRL-----------------EEEKVRAKWRE-IDARQDKLTVDG  3166

Qy  1738  KV-GREEVGVLFQDRNNLR--VKSVSVSQSVSKSEGFNTPALLGTSNSAAMSMERNIG  1793
          :: | :: |: :  :: |    :|  :   | |  | ::| :|| |:: |:   ||::|
Db  3167  LLKGLKESGI-NDINGLEHLFLHPVDTVHELGKI--LTHPTLLQLGESAVQELLNKVS  3222

Qy  1794  TIN--FKYGQDQNTPR--------RFTL-EGGIAQANPQVASALTDLKKEGLE  1835
          :::   | |  | |         ||:| ||| |   :|:|||| | ||:  |:
Db  3223  RMSEALLVGGDHAQQLGEDLGSVIADVGFALAAAGTAKAAEILGEAGISLSKDVLE  3279
```

Query Match 2.68%; Score 247; DB 2; Length 3442;
Best Local Similarity 18.58%; Pred. No. 4.4e-05;
Matches 412; Conservative 292; Mismatches 815; Indels 708; Gaps 103;

```
Qy   948 QTGGMT-------------------------GAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAY  988
             ||:||                               |::|    ||      :||:||
Db  2268 ATGNVTITHGDTIQYTSQDSHTKRSGLLNSRTTTHADQQQTQAIGSTLSADKVFVKGNN 2327

Qy   989 ATQHGWQGREGLKPLYEMQGALIKQLDAHNVR-----HNAPQPDLQSKLETLDLGEHGAEL 1044
             :|:||    |     ||| :: :   || :       |:  |:|
Db  2328 ITVTGSHVVSDAGTYMQAEHDLTLQAATHTTQSTYSHHTKQRGL-------IRNGGASL 2379

Qy  1045 -LNDMKRFRDELEQSATRSVTVLGQHQG------------VLKSNGEINSEFKPSP 1087
            |  | :  ||| ||    : |  |                 ||:|    ||||:|
Db  2380 TLGNQSQRTDSTTTATTTTGSLIGATNGNVTLLAGGHYQQIGSDVLSPHGDIDIHAKKVD 2439

Qy  1088 GKALVQSFNVRSGQDL------------------SKSLQQAVHATPPSAESKLQSMLGH 1129
              :||::|| |:||                     ||  ||  :: :||   |:||:|
Db  2440 ---IIQAHHTSQTTQHTATRQSGLTVALSTPLIAGAQTAQQMQHAAAARSGDPRLQALAGL 2496

Qy  1130 FVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPD 1189
             |:|    :    ||  | :||||    : :| | ||  |  :|   | | |  ||::|
Db  2497 TTALGA-----KNTIDAVRQ-DPRALGGLNAS---------LTVG-------------- 2526

Qy  1190 ADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVN 1249
             |:   :|      |:|:  :        |    | |  :|   :  ||  :|| |: ||
Db  2527 ----RSTHDSTTTTTTSTAAGSNVTAGGNVH---ISATGDGTASTLTIQGSDVRG-- 2574

Qy  1250 LTTRTVLESQGSAELAKKLKNTLLSLDSGGESMSFSRSYG-------GGVSTVFVPTLS 1300
             | : :||||:|:| |   :|||  |||   ::  | |          ||:|  |  |
Db  2575 -DTMTYLKADGDIALL-AAQNTV----TNQRDNRGRSAGVGVAVNLGSGGTSAGLTAHAS 2628

Qy  1301 KKVPVPVPIPGAGITLDRAYNLSFSRTSGG--LNVSFGRD-----------GGVSGNI 1344
              |        ||:|||||:||| |||: |     | :| |               || ||
Db  2629 TST------GSGHSTDLTW--SNSHVGGGNLLAIDAGGDLLMKGAIGTAKHVIADIAGNL 2680

Qy  1345 MVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTE 1404
             || | | | | |  |:||  || |   ||:|   | ||  ||| || : : ||  ||:|
Db  2681 RIESLQDTHQYRSTDRSLGGS------------LTAGAGFSGSSANLNHQTIRSDYASV 2726

Qy  1405 DELPGF----------IHGLTH--GTLTP-AELLQKGIEBHQMKQGSKLTFSVDTSAN-- 1448
            | ||            :| |:    | |:    |:| |  :| |  |  |||||:||
Db  2727 TEQSGLFTGDGGYQLTVGGQTHLIGGAITSNSTAIHNGL-NSLDTGTLILQNIENHANYT 2785

Qy  1449 ---LDLRAGINLNE-------------------DGSKPNGVTARVS------- 1472
               |||| ::| |                      ||| |||| |:|
Db  2786 ATQVNLGGGYSRNGGTVGTDQQGHAATATQVPGTTLPSHNGLSASPPSAMTARDSSHSTT 2845

Qy  1473 -AGLSASANLAAGSRERSTTSGQFGSTTSASNNRP-------------G   66
             ||:|  |  |||  |:| |||  |||  ||::|                |
Db  2846 YSGISQGALTIRDDTAQHALTGHTAAETIATLNRDLLTDTTATSNALTPFDEQRINAGFD 2905

Qy  1507 ----------TFLNGVGAGANL----TAALGVAH-------------SSTHEG 1532
                      ||||  | | | |   | | |||               : :|
Db  2906 IVSGLQRETGTFINNRAAEADLKTRQATAADHAAHDPSNGFNDQQRHALREQAIALTNEA 2965

Qy  1533 KPV--GIFPAFTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHF---- 1586
            |||    || | |::||     |||   || |:| ||  ||| ||  |  ||||
Db  2966 HAIKDAWGPGGTYRQITTALAAG--ASGNVS--------AASSNVSAASSDLAKIMIVNY 3015

Qy  1587 --KDSATTKMLAALKELDDAKPAE-QLHIL-------------QQHFSA-------- 1619
              |||||:| || |::| |||| |  || |                |||:|
Db  3016 VQQQGATAIGHWVATGQLTEGSPLHAALHALLACAGAAASQQHCSSGAQCAAASSVLTGL 3075

Qy  1620 -------KDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGI-- 1671
                    |||| |||  || : ||  ::|:  | |: | ||   |   |  | ||:
Db  3076 FSDPRPEDTTQDR--EAKRNLITSIVTGIASTGNT-DAATATHAA----IAAVDNNWLAA 3128

Qy  1672 ---VELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLRE 1728
                : |    |  || | | | ||        ||| :||||      :|::| |    |
Db  3129 KQYVQMLNEEFEAA----TEKEKGRL--------EEEKVRAKWRE-IDA 3164

Qy  1729 QTEKAILDGKV-GREEVGVLFQDRNNLR--VKSVSVSQSVSKSEGFNTPALLGTSNSA 1784
               |:|   |||   |  | :|| |::|    ||: |:   | ||||  ||  | |||
Db  3165 RQDKLTVDGLLKGLKESGI--NDINGLEHLFLHPVDTVHELGKI--LTHPTLLLQLGESA 3220

Qy  1785 AMSMERNIGTIN--FKYGGQDNTPRRFTLEGG-----IAQANPQVASALT---------D 1828
             |:|  :| |:|   |||  : |  |   |  |     |    | ||||
Db  3221 VQELLNKVSRMSEALLVGGDQHA-QQFGEDLGSVIADVGVALAAAGTFKAAEILGEAGIN 3279

Qy  1829 LKKEGLE 1835
            |: ||||
Db  3280 LSKDVLE 3286
```

```
RESULT     8
Q9K0T0
ID   Q9K0T0          PRELIMINARY;          PRT;   2703 AA.
AC   Q9K0T0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HEMAGGLUTININ/HEMOLYSIN-RELATED PROTEIN.
GN   NMB0493.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002405; AAF40927.1; -.
DR   TIGR; NMB0493; -.
SQ   SEQUENCE   2703 AA;   288263 MW;   214537D1F261F00A CRC64;
```

```
Query Match             2.6%;  Score 245; DB 2;  Length 2703;
Best Local Similarity  19.0%;  Pred. No. 3.9e-05;
Matches  414;  Conservative  282;  Mismatches 766;  Indels 712;  Gaps 105;
```

```
Qy    10 HKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQPSTAAD---G   66
            ||||:|   :||  |  || ||| :: :|:|| |:|| : || ::|: ||||
Db   573 HDLAVNTQTAKNSGH-LLTQTGKIDNRELHNAGEIAA----NNLTLIHSGRLSNDKKGN  626

Qy    67 ISAAHQQKKSFSLR------------GCLGTKKFSFSRSAPQGQPGTTHSKGATLRDLLARDDGE  117
            |||| ||| ||  |             |   |:||     ||  ||   : ||:||||::
Db   627 IRAAHLQLDTAGLHNAGNIILADSGTVTTKNNLRNTGKVSVARLNTEGQTLDNTRGRIEAE  686

Qy   118 TQHEAAAPDAARLTRSGG----------VKRRNMDDMAGRPM------SGE  156
            |::| |  |||||:|||            |  || | |          |||
Db   687 TVN----IQSQLTNQSGHITATEQLTINSRNVDNQNGKLLSANQAQLAVSDGLYNQHGE  742

Qy   157 DKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGST  216
            : :  | || | |  |:|   | ||  ||   |  | |:|  | : |  |  :|
Db   743 IATNRQLSIHDKN------QNTLALNNADGTIQSAGN---------VSLQAKSLA  782

Qy   217 SKATTAHADRVEIAQEDDDSEFQQLHQQRLARERENPPQPPKLGVATPISARFQPKLTAV  276
            | |  | |   |:|:     ||   |          |||  || |  ||: |  |||
Db   783 NNGTLTAGNKDIALTDD-----------FVVERD----LTAGKQLNLS--  816

Qy   277 AESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGK  336
            |  :||  ||: |  |     || |  |  |   :|  :    | |   |  | | :||
Db   817 IKGRLKNTHLQ---------AGHT---LKLNAGNI---DN-----QVTGK  847

Qy   337 DTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVS  396
            ||:|  |  :|  | ||| |||:|||  ||:|| | |||:|||| |||| ||| ||   |
Db   848 -------IIGGEQTDITSEQHVDNRG-------LINS------DGLTHIGAGQTL  882

Qy   397 VDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPGGE-SIRLHDDKI  454
```

```
Db   883  TNTGTGKI-YGNHIALDAQILLNREETTEGSTKAG------AIAARKRLDIGAKEIHNQEG   936
Qy   455  HILHPE-LGVWQSADKDTHSQ------------LSRQADGKLYALKDNRTLQNLSDNKS    500
Db   937  ALLSSEGIFAVGNRLDEQHHAAGMADTFVNGSAGLEVQGD---ALMSVRNMQNINNHFK    992
Qy   501  SE----KLVDKIKSYSV------DQRGQVAILTDTPGRHKMSIMPSLDASPESHISLS-L   549
Db   993  TETYLAKAEKQVRDYTVLGQNTYYQAGKDGLFDNSQGQ-KDQTTATFHLKNGSRIEANQW  1051
Qy   550  HFADAH----QGLLH----------GKSELEAQSVAISHGRLVVAD-------         581
Db  1052  HVRDYHIETYKERIIENRPAHITVGGDLTASGQNWLNKDSRIVVGGRIITDDLNQKEITN  1111
Qy   582  --SEGKLFSAAIPKQGD---------GNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQ   630
Db  1112  QSTTGKGRTDAVGTQWDSVTKKGWYSGRKRQRRTERNHTP---YHDTQL-FTHDFDTP   1165
Qy   631  LNALVKN---NFRQQHAC----PLGNDHQFHPG-WNLTDALVIDNQLGLHHTN         675
Db  1166  VSVIQQNAASPSFQPAASAIKLIDGVSTAAVNGQRIHTGNVVSLNNATVTLPNSSLYTTH  1225
Qy   676  PEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKR   735
Db  1226  PD----NKGW------LVETDPQFADY--RRWLGSDYMLQQLQLDTNHLHKRLGDGYYEQ  1273
Qy   736  LNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKG   795
Db  1274  KLVN------EQIHQLTGYRR------LDGYRSDEE------QFKALMDNG         1306
Qy   796  --DIRSFQIKPG----TQQ---------LERPAQTLSREGISGELKDIHVDHKQNLYAL    839
Db  1307  LTAAKTFGLTPGIALSAEQVARLTSDIVWMENQTVTLS---DGSTQTVLV--PKVYAL   1359
Qy   840  THEGEV-------FHQPREAWQNGAESSS----WHKLALPQSES--------         872
Db  1360  ARKGDLNTSGGLISAEQVLLKLQNGNLTNSGTIAGRQAVLIQARNINSNGNIQADQIGLK  1419
Qy   873  KLKSLDMSHEHKPIATPEDGSQHQ-----------LKAGGWHAYAAPER------G      911
Db  1420  AEKSINI-----------DGGQVQAGRLLTAQAQNINLNGTTQTSGNERNGNTAIDRMAG  1468
Qy   912  PLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSE--RI   968
Db  1469  INVVG-SHTEQVDNRTSDGI--------LSLHASNDINLNAATVSNQVKDGTTQI      1514
Qy   969  RAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDL  1028
Db  1515  TA-GNNLNLGTIRTEHREAYGT--------------LDDENHRHVRQSTEV          1550
Qy  1029  QSKLETLDLGEHGAELL--NDMKRFRDELEQ-----SATRSVT---------VL       1066
Db  1551  GSSIRT---QNGALLRAGNDLKIRQGELEAEEGKTVLAAGRDVTISEGRQITELDTSVS   1606
Qy  1067  GQHQGVLKSNGEINSEPKPS---------PGKALV--QSFNVN------RSG        1101
Db  1607  GKSKGILSST-KTHDRYRFSHDEAVGSNIGGGKMIVAAGQDINVRGSNLISDKGIVLKAG  1665
Qy  1102  QDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQR--DPNDKTALT  1159
Db  1666  HDIDISTAHNRYTGNEYHESKKSGVMG--TGGLGFT-------IGNRKTTDDTDRTNIV   1715
Qy  1160  KSRLIL----DTVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVK   1214
Db  1716  HTGSIIGSLNGDTVTVAGNRYRQTGSTVSSPEGRNTVTAKSIDVEF--ANNRYATDYAH   1772
Qy  1215  HYTDMGFT---HNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNT  1271
Db  1773  TQEQKGLTVALNVPVVQAAQNFIQAAQNVGKSKNKRVNAMAAANAAWQ-SYQATQOMQQF  1831
Qy  1272  LLSLDSGESMSFSRSYGGGVSTVFVPTLSK-------KVPVPVIPGAGITLDRAYNLSF   1323
```

```
Db  1832  APSSSAGQGQNNNQSPSISVSITYGEQKSRNEQKRHYTEAAASQIIGKQT-------TL   1884
Qy  1324  SRTSGG----LNVSFGRDGGVSGNIMVATGHDVM--PYMTGKKTSAGNASDW-LSAKHKI  1376
Db  1885  AATGSGEQSNINITGSDVIGHAGTALIADNHIRLQSAKQDGSEQSKNKSSGWNAGVAVKI  1944
Qy  1377  SPDLRIGAAVSGTL-QGTLQNSLKFKLTEDELFGFIHGLTH-----------GTLT    1420
Db  1945  GNGIRFGITAGGNIGKGKEQGG---------STTHRHTHVGSTVGKTTIRSGGDTTLK   1993
Qy  1421  PAELLQKGIE------------------------HQMKQ-----GSKLT----     1440
Db  1994  GVQLIGKGIQADTRNLHIESVQDTETYQSKQQNGNVQVTVGYFSASGSYRQSKVKADHA   2053
Qy  1441  ----------FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSR     1486
Db  2054  SVTGQSGIYAGEDGYQIKVRDNTDLKGGIITSSQSAEDRGKNLFQTATLTAS-DIQNHSR  2112
Qy  1487  ERSTTSGQFGS-----------TTSASNNRPTFLNGVGAG---------ANLTAALGV----AH  1526
Db  2113  YEGRSFGIGGSFDLNGGWDGTVTDKQGRPTDRISPAAGYGSDGDSKNSTTRSGVNTHNIH  2172
Qy  1527  SSTHEGKPVGIFPPAFTSTNVSAALALDNRTSQSISLELKRA--EPVTSNDIS--ELTST  1581
Db  2173  ITDEAGQLARTGRTAKETEARIYTGIDTETADQHSGHLKNSFDKDAVAKEINLQREVTKE  2232
Qy  1582  LGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQ  1641
Db  2233  ERYQEARTL--LEAEL                                             2268
Qy  1642  QAADSHSME-----------------------------LGSASHSTTYNNLSRINN     1668
Db  2269  QNTDSEAEKAAFRASLGQVNAYLAENQSRYDTWKEGGIGRSILHGAAGGLTTGSLGGILA  2328
Qy  1669  DGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLSQTPFSSASVSMELKDGLRE  1728
Db  2329  GGGTSLAAPYLDKA------AENLG----PAGKAAVNALGGAAIGYAT--------     2366
Qy  1729  QTEKAILDGKVGREEVGVLFQDRNN--LRVKSVSVS---------QSVSKSEGFNTPALLL  1778
Db  2367  -----------GGSGGAVVGANV-DWNNRQLHPKEMALADKYAEALKREVEKREGRKI------  2412
Qy  1779  GTSNSAAMSMERNI  1792
Db  2413  -SSQEAAMRIRRQI  2425
```

RESULT   9
Q14789
ID   Q14789   PRELIMINARY;   PRT;   3259 AA.
AC   Q14789; Q14398;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   GIANTIN (GCP372) (MACROGOLGIN) (GOLGI AUTOANTIGEN, GOLGIN SUBFAMILY B,
DE   1).
GN   GOLGB1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=009606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94187728; PubMed=7511208;
RA   Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA   Renz M.;
RT   "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
RT   protein (giantin) [retracted in Mol Cell Biol 1995 Jan;15(1):591].";
RL   Mol. Cell. Biol. 14:2564-2576(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94257116; PubMed=8198703;
RA   Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,

```
RA   Renz M.;
RT   "Macrogolgin--a new 376 kD Golgi complex outer membrane protein as
RT   target of antibodies in patients with rheumatic diseases and HIV
RT   infections.";
RL   J. Autoimmun. 7:67-91(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95109974; PubMed=7802676;
RA   Sohda M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
RT   "Molecular cloning and sequence analysis of a human 372-kDa protein
RT   localized in the Golgi complex.";
RL   Biochem. Biophys. Res. Commun. 205:1399-1408(1994).
CC   -!- FUNCTION: MAY PARTICIPATE IN FORMING INTERCISTERNAL CROSS-BRIDGES
CC       OF THE GOLGI COMPLEX.
CC   -!- SUBUNIT: DISULFIDE-LINKED HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI MEMBRANE.
CC   -!- ALTERNATIVE PRODUCTS: DIFFERENTS ISOFORMS MAY ARISE BY ALTERNATIVE
CC       SPLICING.
CC   -!- DISEASE: ANTIGEN IN CHRONIC RHEUMATOID ARTHRITIS AND IN THE
CC       AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME.
DR   EMBL; X75304; CAA53052.1; -.
DR   EMBL; D25542; BAA05025.1; -.
DR   MIM; 602500; -.
KW   Golgi stack; Antigen; Coiled coil; Transmembrane;
KW   Alternative splicing.
FT   DOMAIN          1      3235       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     3236      3256       POTENTIAL.
FT   DOMAIN       3257      3259       LUMENAL (POTENTIAL).
FT   DOMAIN         48       110       COILED COIL (POTENTIAL).
FT   DOMAIN        127       223       COILED COIL (POTENTIAL).
FT   DOMAIN        238       448       COILED COIL (POTENTIAL).
FT   DOMAIN        460       526       COILED COIL (POTENTIAL).
FT   DOMAIN        545       593       COILED COIL (POTENTIAL).
FT   DOMAIN        677       956       COILED COIL (POTENTIAL).
FT   DOMAIN        969      1028       COILED COIL (POTENTIAL).
FT   DOMAIN       1062      1128       COILED COIL (POTENTIAL).
FT   DOMAIN       1154      1245       COILED COIL (POTENTIAL).
FT   DOMAIN       1301      1779       COILED COIL (POTENTIAL).
FT   DOMAIN       1828      2781       COILED COIL (POTENTIAL).
FT   DOMAIN       2797      2857       COILED COIL (POTENTIAL).
FT   DOMAIN       2872      2993       COILED COIL (POTENTIAL).
FT   DOMAIN       3026      3102       COILED COIL (POTENTIAL).
FT   DOMAIN       3133      3185       COILED COIL (POTENTIAL).
FT   DOMAIN       2420      2423       POLY-GLU.
FT   DOMAIN       2993      2996       POLY-SER.
FT   VARSPLIC        1        39       MISSING (IN REF. 3).
FT   VARSPLIC      215       215       A -> AQLSSM (IN REF. 3).
FT   CONFLICT     1765      1765       D -> G (IN REF. 3).
FT   CONFLICT     2950      2950       H -> D (IN REF. 3).
SQ   SEQUENCE   3259 AA;  376075 MW;  60376A20D8A178DD CRC64;
```

```
Query Match           2.6%;  Score 244.5;  DB 4;  Length 3259;
Best Local Similarity 18.6%;  Pred. No. 5.7e-05;
Matches  389;  Conservative 311;  Mismatches 816;  Indels 577;  Gaps  92;

Qy      23 GHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQ-PSTRADGISAAHQQKKSFSLR  80
              ::   |  |  |: :      :| |  |            :|:|  |  |  :  |
Db     603 GEGIA------PIKMKVFLEDTGQDFPLMPNEESSLPAVEKEQASTEHQSRTSEEI-  652

Qy      81 GCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQHEAAAPDAAR---------  129
                         :  |:  :  :: :  |   :  ||  | |
Db     653 --------SLNDAGVELKS--TKQDGD-KSLSAVPDIGQCHQDELERLKS  691

Qy     130 ------LTRSGGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNN------FGQMR  175
                        :::|:  :  |    |          ||
Db     692 QILELELNFHKAQEIYEKNLDEKA-----KEISNLNQLIEEFKKNADNNSSAFTALSEER  746

Qy     176 QTMLSKM-------------------AHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSK  218
           |:|:  :                      :  | : | :: :  |::    |     |
Db     747 DQLLSQVKELSMVTELRAQVKQLEMNLAEAERQRRLDYES--QTAHDNLLTEQIHSLSI  803

Qy     219 ATTAHADRVEIAQED-DDSEFQQLHQQRLARERENPPQQPPKLGV-----ATPISARFQP  271
           ::: : ::::  :| |     |   ::| | :  : : ||        : ||  : :|
Db     804 EAKSKDVKIEVLQNELDDVQLQFSEQSTLIRSLQSOLQNKESVLEGAERVRHISSKVEE  863
```

```
Qy     272 KLTAVAESVLEGTDTTQSPLKP------QSMLKGSGAGVTPLAVTLDKGKLQLAPDN--  322
           |::: ||  :|    |  :|          :   |  || | :|  |  ||  ||
Db     864 LSQALSQKELEITKMDQLLEKKRDVETLQQTIEEKDQQVTEISFSMTEKMVQLNEEKFS  923

Qy     323 -------PPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSV  374
                     :  | |: |  |  | ::  ||    |  | |  |  | :  :
Db     924 LGVEIKTLKEQLNLLSRAEEAKKEQVEEDNEVSSGLKQN------YDEMSPAGQISK  974

Qy     375 LHNSHPGEI------------KGKLAQAGTGSVSVDGKSGKIS---LGSGTQSHNKTMLSQP--  421
              : : |            |||  |    |  |  | |:: |     |  | | :| |:
Db     975 EELQHEFDLLKKENEQRKRKLQAALINRKELLQRVSRLEEELANLKDESKKEIPLSETER  1034

Qy     422 GEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADG  481
             |  |  ||   | |   |   | : |    | :   |   | |  |    | |   |
Db    1035 GEVEEDKENKEYSEKCVTSKCQEIEYL--KQTISEKEVEL-QHIRKDLEEKLAAEEQF  1090

Qy     482 KLYALKDNRTLQN---------------------LSDNKSSEKLVDKIKSYSVD--QRGQVAILTDTPGR  528
             :  :|  | :                    :| | | : |  :|   |    :| | |  |  | |
Db    1091 QALVKQMNQTLQDKTNQIDLLQAEISENQA---IIQKLITSNTDASDGDSVALVKET--  1144

Qy     529 HKMSIMPSLDASPESHISLSHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFS  588
             | | |    :    |: |  | |      || ::| |   ::  | |   |: ::
Db    1145 --VVISPPCTGSSE------------------HWKPELEEKILALEKEK--EQLQKKLQE  1182

Qy     589 A---AIPKQGD--GNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN  637
           |      | |     |: |  |   |  | |  |    |  |   | | :|
Db    1183 ALTSRKAIIKKAQEKERHLREELKQQKDDYNRLQEOFDEQSKENENIGDQLRQLQIQVRE  1242

Qy     638 NF------RQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHLGSL  690
                    | : :       || ::   ::|| |   |    : |     | :  |
Db    1243 SIDGKLPSTDQQESC------SSTPG--LEEPLFKATE--QHHTQP----VLE------  1281

Qy     691 ALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAY--LLKDGEVKR---LNINQSTSS  744
           |  |  |:||     |   |  |  |: |    |      | ||     | :| || |
Db    1282 ------------SNLCPDWPSHSEDASALQGGTSVAQIKAQLKEIEAEKVELELKVSSTTSE  1331

Qy     745 IKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKP  804
           |  |  | :| |  |          |  | |   :      || | | |:  |  |:|
Db    1332 LTKKSEEVFQLQEQINK--------QGL--------------EIES--LKT  1358

Qy     805 GTQOLERPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFHQPRE--AWONGAESSS  861
             |   |   |  | ||  |  |   |  ||  | |    | |:| |    |: ||| |
Db    1359 VSHEAEVHAESLQQKLESSQLQIAGLEHLRELQPKLDELQKLISKKEDVSYLSGQLSEK  1418

Qy     862 WHKLA-------LPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVG  916
           | ||           |  | |||: |  :  |  | | | ||   | |  |  |  | |
Db    1419 EAALTKIQTEIIEQEDLIKALHTQLEMQ--AKEHDERIKQLQVELCEMKQKPEE----IG  1472

Qy     917 TSGSQTVFNRLMOGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERI--RAYAF  973
              :|  :::         | | |  : :: |   | |   |   |  |:  :   | |
Db    1473 EE------SRAKQQIQRKL-----QAALISRKEALKENKSLQEELSLARGTIE  1514

Qy     974 NPTMSTPRPIKNAAYATQHGWQGRE------GLKPLYEMQGALIKQLDAHNVRHNAPQPD  1027
           | |:| |      |: | || ||         |:|       |:  ||  |  :|  | |
Db    1515 RLTFKSL------ADVESQVSAQNKEKDTVLGRLALLQEERDKLITEMD--RSLLENQS  1564

Qy    1028 LQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSP  1087
            ||| |  ||  |  ||  |   |  |||  :  | | | |  | | |  |  | |  |
Db    1565 LSSSCESLKLALEG--LTEDKEKLVKEIESLKSSKIAESTEWQ--EKHKELQKEY----  1615

Qy    1088 GKALVQSFNVNRSGQDLSKSLQQAVHATPSAESKLQSMLGHFVSAGVDMSH----QKG  1142
           |  |:| |   |   | ||||  | |  |  |:| |:|    : ||| |:       ||
Db    1616 -EILLQSY------ENVSNEAERIQHVEAVRQEK-QELYGKLRSTEANKKETEKQLEA  1667

Qy    1143 EIPLGRQRDPNDKTALTKSRLILDTV-----TIGELHELADKAK----LVSDHKPDADQ  1192
           |:|    |  ||| ||:|:|||| |         |  |  |  |       | |  | |
Db    1668 EQEMEEMEKMRKFAKSKQQKILELEEENDRLRAEVHPAGDTAKECMETLLSSNASMKEE  1727

Qy    1193 IKQLRQOFDTLREK----------RYESNPVKHYTDMGFTHNKALEANYDAVKAFINAF  1241
           ::::: :  : :  :          | :   ::  |:: | | |  | |   | | | :
Db     747 DQLLSQVKELSMVTELRAQVKQLEMNLAEAERQRRLDYES--QTAHDNLLTEQIHSLSI  803
```

```
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031128; CAA20006.1; -.
DR   FLYBASE; FBgn0025392; EG:49E4.1.
SQ   SEQUENCE   5327 AA;  575942 MW;  FEFEE23A118FF38A CRC64;

Query Match             2.5%;  Score 240.5;  DB 5;  Length 5327;
Best Local Similarity   18.4%;  Pred. No. 0.00022;
Matches  370;  Conservative  271;  Mismatches  803;  Indels  567;  Gaps  81;

Qy     18 AHNPVGHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQPSTAADGI------SAAH  71
Db   3517 ASRPTSVAESVKDEAEKSKEESRRESVAEKSSLASK- EASRPASVAESVKDEAEKSKEE  3574

Qy     72 QQKKSFSLRGCLGTKFSRSAPQGQ-PGTTHSKGATLRDLLARDDGETQHEAAAPDAA  128
Db   3575 SRRESVAEKSPLASKEASRPASVAESVKDEAEKSKEVSRRESVAEKSPLPSKEASRPTSV  3634

Qy    129 ------------RLTRSGGVKRRNMDDM-AGRPMVKGGSGEDKVPTQQKRHQLNNFGQM  174
Db   3635 AESVKDEADKSKEESRRESGAEKSPLASMEASRPTSVAESVKDETEKSKESRRESVTEK  3694

Qy    175 RQTMLSKMAHPASA- NAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQE  232
Db   3695 SPLPSKEASRPTSVAESVKDEAEK-------SKEESRRESVAEKSPLASKESSRPASVAE  3747

Qy    233 DDDSEFQQLHQQRLARERENPPQPPKL-- GVATPISARFQPKLTAVAESVLEGTDTTQ  288
Db   3748 SIKDAEGTKQE-- SRRESMEPSGKAESIKGDQSSLASKETSRPDSVVESVKDETE-  3801

Qy    289 SPLKPQ-SMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHAS  347
Db   3802 ---KPEGSAIDKSQVASRPESVAVS------AKDEKSPLHSRPESVADKSPD------AS  3846

Qy    348 SDGSQHLLLDNKGHLFDIKSTATS------YSVLHNSHP---GEIKGKLAQAGTGSVSV  397
Db   3847 KEASRSL--------SVAETASSPIEEGPRSIADLSLPLNLTGEAKGKLP---TLSSPI  3894

Qy    398 DGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHIL  457
Db   3895 DVAEGDFLEVKAESSPRPAVLSKPAEFSQPDTGHTASTVDEASPVLEEIEVVEQ----  3949

Qy    458 HPELGVWQSADKDTHSQLSRQADGKLYALKDNR-TLQNLSDNKSSEKLVDKIKSYSVDQ  515
Db   3950 HTTSGV-------GATGATAETDLLDLTETKSETVTKQSETTLFETLTSKVES------  3995

Qy    516 RGQVAILTDTPGHRKMSIMPSLDASPESHISLSLHFA---DAHQGLLHGKSE-LEAAQ  568
Db   3996 ------KVEVLESSVKQVEEKVQTSVKQAETTVTDSLEQLTKKSSEQLTEIK------  4041

Qy    569 SVAISHGRLVVADSEGKLFSAAI---PKQGDGNELKMKAMPQHALDEHFGHDHQISGFF  624
Db   4042 SVLDTN----ISNVTNLFSTAVETIEKKVQDVTEKVIEKATEH-VSEHVTTGESSTET  4095

Qy    625 HDDHGQLNALVKNNFRQQHACPLGN------------DHQFHPGWNLTD  661
Db   4096 SQEKSSLDLGTFSELRETHITTVGSPEFTVTICERDEPVLHDIKEEDEEHRFSPPSDVDK  4155

Qy    662 ALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGK----LHYFDQLTKGWTGAESDCKQ  716
Db   4156 AAIIPPQ-PMRPLSPREEEVAKIVADVAKVLKSDKDITOLIPDFDE---------RQ  4202

Qy    717 LKKGLDGAAYLLKDGEVKRLNINQSTSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDK  776
Db   4203 LEEKLKSTADTEEESDKSTRDEKSLEISVKVEIESEKSSPDQKSGPISIEEKDIEQSEK  4262

Qy    777 AQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNL  836
Db   4263 A------------QLRQGILASSRP-----ESVASQPESVPSPSQS--  4291

Qy    837 YALTHEGEVFHQPREAWQNGAESSSWHKLALPQS-ESKLKSLDMSHEHKPIATFEDGSQH  895
Db   4292 -AASHE---HKEVEL---SESHKAEKSSRPESVASQVSEKDMKTSRPASST----SQF  4338
```

```
Db   1728 LERVKMEYETLSKKFQSLMSEKDSLSEVQDLKHQIE----DNVSKQANLEAT-------  1776
Qy   1242 KKEHHGVNLTTRTVLESQGSAELAKKLKNTLLSLDS----GESMSFSRSYGGGVSTVF-  1295
Db   1777 EKHDNQTNVTEEGTQSIPGETE---EQDSLSMSTRPTCSESVPSAKSANPAVSKDFSS  1831

Qy   1296 --------VPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGG----LNVSFGRDGGV  1340
Db   1832 HDEINNYLQQIDQLKERI------AGLEEEKQKNKEFSQTLENEKNTLLSQISTKDGEL  1884

Qy   1341 ------SGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQG  1392
Db   1885 KMLQEEVTKMNLLNQQIQEELSRVTKLKETAEEEKD------------DLEE  1924

Qy   1393 TLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLLFSVD--------  1444
Db   1925 RLMNQLA---ELNGSI-GNYCQDVTDAQIKNELLESEMKNLKKCVSELEEEKQQLVK  1977

Qy   1445 --TSANLDLR------AGINLNEDGSKPNGV---------------TARVSAGLS  1476
Db   1978 EKTKVESEIRKEYLEKIQGAQ-KEPGNKSHAKELQELLKEKQQEVKQLQKDCIRYQEKIS  2036

Qy   1477 ASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVG  1536
Db   2037 A--------LERTVKALEFVQTESQKDLEITKEN-------LAQAVEHRKKAQAELAS  2079

Qy   1537 IFPAFTSTNVSAALAL-DN-------RTSQSISLELKRAEPVTSNDIS-ELTSTLGKHF  1586
Db   2080 FKVLLDDTQSEAARVLADNLKLKKELQSNKESVKSQMKQKD----EDLERRLEQAEEKHL  2135

Qy   1587 KDSATTKMLAALKELDDAKPAEQLHILQOHFSAKDVVGDERYEAVRNLKKLVIRQQAADS  1646
Db   2136 KEKKN------MQEKLDALRREKVHL------EETIGIQVTLNKKDKEVQQLQENLDS  2182

Qy   1647 HSMELG--SASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDI  1704
Db   2183 TVTQLAAFTKSMSSLQDDRDRV-IDEAKKWERKFSDAIQSKEEEIRLKE--DNCSVLKDQ  2239

Qy   1705 IKQLQSTPFSSASVSM-ELKDGL-REQTEKAILDGKVGRE------------  1742
Db   2240 LRQM------SIHMEELKINISRLEHDKQIWESKAQTEVQLQQKVCDTLQGENKELLS  2291

Qy   1743 ------EVGVLFQDRNNLRYKSVSVSQSV5KSEGFNTPALLGTSNSAAMSM  1788
Db   2292 QLEETRHLYHSSQNELAKLESELKSLKDQLTDLSNSLEKCKEQK------GNLEGIIRQQ  2345

Qy   1789 ERNIGTINFKYGQ--DQNTPRRFT-LEGGIAQANPQVASALTDLKKEGLEM  1836
Db   2346 EADIQNSKFSYEQLETDLQASRELTSRLHEEINMKEQKIISLLSG-KEEAIQV  2397
```

```
RESULT  10
076891

ID   076891;    PRELIMINARY;    PRT;   5327 AA.
AC   076891;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   EG:49E4.1 PROTEIN.
GN   EG:49E4.1.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Papagiannakis G., Spanos L., Siden-Kiamos I., Louis C.;
RT   "sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
```

```
RESULT  11
Q9NJ17
ID   Q9NJ17        PRELIMINARY;     PRT;   5476 AA.
AC   Q9NJ17;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   SPLIT ENDS.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Laverty T.,
RA   Suh C., Voas M., Williams A., Rubin G.M.;
RT   "A Genetic Screen for Novel Components of the Ras/Mitogen-Activated
RT   Protein Kinase Signaling Pathway That Interact With the yan Gene of
RT   Drosophila Identifies split ends, a New RNA Recognition Motif-
RT   Containing Protein.";
RL   Genetics 154:695-712(2000).
DR   EMBL; AF184612; AAF26299.1; -.
SQ   SEQUENCE  5476 AA;  590531 MW;  93FAA8C7860770C2 CRC64;

Query Match           2.5%;
Best Local Similarity 19.5%;           Score 240.5;  DB 5;  Length 5476;
Matches  422;  Conservative  256;  Mismatches  744;  Indels  739;  Gaps  106;

Qy       6 LGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAAS------------KNRGKMP   55
             |   :::   |                                        || |  :
Db      25 LVTQYQANNSTSLAN-----------------------SNSSPSVSASVFATAAGGSSERSNRDRPY   72

Qy      56 RIHQPSTAADGISAAHQQKKSF-------SLRGCLGTKK-FSRSAPQQGPGTTHSKGATL  107
            ::| |:|| |  ::  :: |                   |   :   |   ::  :||
Db      73 RNGSASVQGGGINSSNTTTTAACTAGGSGGAIGTGTGGLVGSGPGGVP-------QAL  125

Qy     108 RDLLARDDGETQHEAA--AP----DAARLTRSGGVKRRNMDMAGRPMVKGGSGEDKVP  160
               :::  |   |        |    ::::                          | |
Db     126 GDRSSTQNIHQNHQSARVAPPQSWYEAATAATTAQLKS---------------SGGSG-  168

Qy     161 TQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSP------PHI--PGSHHEIKE  210
                                                            ::|
Db     169 ------------------------------NAG----------ASAAVGFTMSSSPINHHPHQHPHLQNPQHPHYTSS  206

Qy     211 EPVG-----------STSKATTAHADRVEIAQEDDD------SEFQQLHQQRLAR-  248
            :  |                ::  |       |:::
Db     207 PVVGAGSCPSAAQGQPQIQSQSQTTAVHRSVAYAGSAADDLLNTATSRNMLLHSSKLNKL  266

Qy     249 ---------ERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQ--  294
                                |:|  :                 :
Db     267 LKGAGATGSGERSGSESPGRAGGATPLTTTSTITTNNSFSSNSLNNTITTATPTMPTIAS  326

Qy     295 ---SMLKGSG--AGV-----------TPLAVTLDKGKLQLAPDNPPALNTLLKQT  333
              ::   ::  : |               :|::|     |:   :  |  ::||:
Db     327 GAAGSVGLGSGAEAGVCSNSCTASGDILNVAVLAAAVDNG----VPTHPIRTRHNLH--  380

Qy     334 LGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQGTG  393
             | | :||  |  ||  |::::||: ::::  | |  ::  ::  |:  |    :
Db     381 -GRSTTSSSRSHSRSPSSY----SSHSSSSSSHSSSHS--HASSPVQSSGNCAMA-----  429

Qy     394 SVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDK  453
            :    :::  ::  |:        ::|  |  | |
Db     430 -------------EGRSSR-TVNSVTVTSNSS-----------------NPSGTA-  453

Qy     454 IHILHPELGVWQSADKDTHSQLSRQADGK------------QNLSDNKSSE  502
             ::  |  |   :  :   :|::    |                : : |:
Db     454 --VTVSSAGVGGGCGSSSSSSSSSSSSGGSCLTANPVVHSEDNRPLAIRVRNLPARSSDT  511

Qy     503 KLVDKIKSYSVDQRGQVA-----ILTDTPGRHKM-------SIMPSLDASPESHISLSLHFA  552
            ::   :::: ::|| |        :| | : ::          :|    |   | | |
```

```
Qy     896 QLKAGGWHAYAAPERGPLAVGTGSGQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGA   955
             |                   :::|         |  |       |       ::|
Db    4339 STKEGDEETTES-------LLHSLTTTETVETKQME--------------EKSSFESVSTSVT  4380

Qy     956 EGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLD  1015
            |  |  :  ::|  ::| |: ::                   :| |  |
Db    4381 KSTVLSSQSTVQLREESTSESLSSSLKVEDSS-------RRESLSSLLAEKGGI------  4427

Qy    1016 AHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKS  1075
            | |  |           |   | |  |:     |          :::   : |  |
Db    4428 ATNTSLKEDTSASASQLE--------ELL------VQSEECSSESIVSEIQTSIAQKS  4471

Qy    1076 NGEI---------------NSEFKPSPGKALVQ-----------SFNVNRSGQD-  1103
                              ::                       :
Db    4472 NKEIKDARETKVTSQFTTTTSSATKDDSLKETVAEFLATEKIVSAKEAFSTEATKSADDC  4531

Qy    1104 LSKSLQQAVHATPPSA-----------ESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDK  1155
            ::    :   ::            ||::|:  |  |:                |: |:|
Db    4532 LKKTTASAVSSTSASQRALFVGTDESRRESLLSQASESRLTHSDPEDEEPAD---DVDER  4588

Qy    1156 TALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKH  1215
            |  |:     : |  :         :::                    :     :|:
Db    4589 SSVKESR-----SKSIATI------MMTSIYKPSEDM---------------EPI--  4617

Qy    1216 YTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLES--QGSAELAKKLKN--  1270
                                              :  :  :|   | :  | |:|
Db    4618 -----------SKLVEEEHEHVE--ELAQEVTSTSKTT-TLLQSSEQSSTTTSSTSKTGAS  4664

Qy    1271 ----TLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSR  1325
                ::  :::|:|:::  :|| |:   :  |  |  |:  |  :| ||    |  | |
Db    4665 RVESITLTQMD----QQTSQSGQDPADRKTPPT----APVSPGV------  4700

Qy    1326 TSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAA  1385
             :|  :  || | ||            :  |  |               |   :  |:
Db    4701 ----------KAMSSTG------------SAGSV-------------------IGAG  4716

Qy    1386 VSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGCTLTPAELLQKGIEHQMKQGSKLIFSVDT  1445
            :|| |||:  ||  ||  |  ||: |:| |  |  | :            | | :::  |
Db    4717 AGAVAAGGKCESSAASIVS--------SSGPMSP------KDISGKSSPGALTSESQSI  4761

Qy    1446 SANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTSASNNR  1505
            |   |  |:                 : :         :: |:     |:
Db    4762 PTPLGRESHTDTPESSPKPTSPFPRVSKDELKSLEMQHHSQEQMLAGAAAAAGAECEGDI  4821

Qy    1506 PTF--LNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNR---TSQSI  1560
                |  |     |  |  |  :     |             :    |       :|::
Db    4822 PELHELRGL---ECTTALS---GSTDK------IITTTTTVTKVISADGKEIVTEQKT  4868

Qy    1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKEL--------DD----AKP  1606
            | ::| ::   |: |   | || |: ||  |   :   |         |      | |
Db    4869 VTTTDSSEPDSEKVVVTTTRTTSESERDQLLPKEVALLRGLIYRASTPGSEDDEDLLLGSP  4928

Qy    1607 AEQLHILQQHFSA-----------KDVVGDERYEAVRNLKKLVIRQQARDSHSMELGSASHSTT  1659
            |     ::|:               : | |   :::  | : ::  ::: |  ::  | |
Db    4929 RSATSYELQHSSSGVSKRSDLDADGDE--------SQDDIPQYGSEEHSTA  4972

Qy    1660 YNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPF-----  1713
             ::  | |    :|||: ||            |                 |
Db    4973 RSILLPRTADPMA---TSFYGALPDSF-------DVVMKPSTEPIPIQGAP  5013

Qy    1714 ---SSASVSMELKDGLREQT-----------EKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQ  1763
                 | | ::| |                |   :  |       ::|  |  ::|:
Db    5014 SGDSQSSESVESS-----SQTWAGHKFLDQADKDFQRALEE--HVQARGAEVMSSVTAKY  5066

Qy    1764 SVSKSEGFNTPALLLGTSN----------SAAMSMERNIGTINFKYGQDQNTPRRFTLEGG  1814
            |:   | |  || |:||             | :| |:  |     | :          ||
Db    5067 SYSPSKAEEMEQIVSGTAERQRPLSDVQRARVAESGFATVGSVASQQQQQEK---GGE  5122

Qy    1815 IAQANP----------QVASALTDLKKEGLE  1835
                              |:| | |:|  ||:
Db    5123 VEQAVPTTTAVTASTTATASSTGALPKDRLE  5153
```

Qy 1475 ------------LSASANLAAG--SRERSTTSQFGS--- 1510
Db 1417 SGHTPSSATPSWQLSMHMNLSTGLNSHQTSSASGNSNSSSGTVSSSASNSRHKFLDIDEL 1476

Qy 1511 ------------------------GVGAGANL-TAALGVAHSSTHEGK 1533
Db 1477 QPSDIVKSVLAKKSVFDDDFQRLNKNQWDPSSSDFALGSSSNIVTGSSLVANVSRHPGG 1536

Qy 1534 PVGIFPAFTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTK 1593
Db 1537 P-----CSGNTSPALP------NLAATKATPIIGNCSGGLGNSTGS--KSAGLLQ 1578

Qy 1594 MLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQADSHSMELGS 1653
Db 1579 RLSSLSPMN--SPQASMSPYNSPSPSVGG-----VTACLGQLTKPAAPGTASAGL-- 1628

Qy 1654 ASHSTTYNNLSRINNDGIVELLHKHFDAALP-ASSAKRLGEMMNNDPALKDIIKQLQSTP 1712
Db 1629 -SGGTAASSSPAANSGPTKGLQYPFSHPPLPNTAAPPAVQPAPPLPEMGKQSRLTG 1687

Qy 1713 FSSA---SVSMELKDG-----LREQTEK-AILDGKVGREEVGVLFQDRNNLRVK-SVSVS 1762
Db 1688 QSSGNNLTKSLSVPDGPQSSPARVQLQKSASVPGSTNVGAPSSLSLDSTTASVETSASIS 1747

Qy 1763 QSVSKSEGFNTPALL-----------LGTSNS-AAMSMERNIGTINFKYGQDQ 1803
Db 1748 SSTSNGNSSLTSAAIHVQKPQQSTFVEEEHTKKSGTSTSQSSSSSKKISSTHDKLHSKH 1807

Qy 1804 N 1804
Db 1808 N 1808

RESULT 12
Q9U6C3

Query Match   2.5%; Score 240.5; DB 5; Length 5533;
Best Local Similarity   19.5%; Pred. No. 0.00023;
Matches 422; Conservative 256; Mismatches 744; Indels 739; Gaps 106;

Qy 6 LGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAAS-----------LAAEG 55
Db 82 LVTQYQANNSTSLAN--------------SNSSPSSVSASASVFATAAGGSSERSRNRDRPY 129

Qy 56 RIHQPSTAADGISAAHQQKKSF------SLRGCLGTKK--FSRSAPQOQPGTTHSKGATL 107
Db 130 RNGSASVQGGGINSSNTTTTAACTAGGSGSGAIGTGTGGLVGSGPGGVP--------QAL 182

Db 512 SLKDGL-FHEYKKHGKVTWVKVVGQNSERYALVCFKKPDDVEKALEVSHDKHFFGCKIEV 570
Qy 553 DAHQGL------LHGKS------ELEAQSVA--ISHGRLVVADSEGKL 586

Db 571 EPYQGYDVEDNEFRPYEAELDEYHPKSTRTLFIGNLEKDITAGELRSH-----FEAFGEI 625
Qy 587 FSAAIPKQG---------DG----NELKM--KAMPQHA--LDEHF 614

Db 626 IEIDIKKQGLNAYAFCQYSDIVSVVKAMRKMDGEHLGSNRIKLGFGKSMPTNCVWID--- 682
Qy 615 GHDHQIS-GFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHH 673

Db 683 GVDEKVSESFLQSQFTRFGAVTKVSIDR------NRQLALVLYDQV---- 722
Qy 674 TNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDC------KQLKKGLDGAAYL 727

Db 723 QNAQA-AVKDM--RGTI-LRRKKLQV-----DFASRECQDAFYDKQEKQQQQSS---- 767
Qy 728 LKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPE--------PGDALQGLNKDDKA- 777

Db 768 ---GSNPRFSRYESSASSLQSRSRASSFSRHQNNSNDDCSPINTPGGASSGISSASNLIN 824
Qy 778 QAMAV----IGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTLSREGISGELKDIHVDHK 833

Db 825 QSTSINISNIGTNACSAM------------PAPSLASAVVS-----CNVNAS 859
Qy 834 QNLYALTHEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGS 893

Db 860 GTVPAST-----------SMPSGVSSSS--SSLPMSPAAL-----AQRHRMVRNARQTV 900
Qy 894 QHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRL--------MQG---------VKGK 934

Db 901 DCDFNEVGRLRFRSSEE---VSGGAGNSTQFEDVRCDSPVTARQGSAVNCFTGPTAAVGE 957
Qy 935 VIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGW 994

Db 958 SIDGTLNNNQITGGAEGFTGSGGSILSRRRCGK---------TPK--- 993
Qy 995 QGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDE 1054

Db 994 ---DLHPVHNQRIQLAEQVE----------ECPSSGDEG--VVSPRKRIKMD 1030
Qy 1055 LEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNV---NRSGQDLSKSLQ 1109

Db 1031 YHHHHHS------NASGVESTGEHSSINKPSP--LLLSNCDVIHDPLNR-----KSEI 1076
Qy 1110 QAVHATPPSAESKLQSMLGHFVSA--GVDMSHQKGEIPLGRQRDPNDKTALTKSRLLIDT 1167

Db 1077 RRVSETPSGSPS--IKFPGHLPSAPQSLMLSCRRPSIDVGA-----LSALSSSSAF-- 1125
Qy 1168 VTIGELHELADAKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKAL 1227

Db 1126 ------RHGIVGASSMDQQHMNASAAAKRRVTTTMQQPSSSS--------TTNSSS 1169
Qy 1228 EANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSY 1287

Db 1170 GSGLGGISSLTPA-DEYHHVSRGRGHQLHSHSHEAS---------GGESADGSRP- 1216
Qy 1288 GGGVSTV-----FVPTLSKKVPVP---VIPGAGITLDRAYNLSFSRTSG 1328

Db 1217 --GTPLCDERPEVLPTEPRRLPPPRERVRERTRDVMWLPLPKFGVLF---FQQQQSRSSG 1271
Qy 1329 GLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSG 1388

Db 1272 G-----GG------AGNSYLQQQLGGGSTGGLGC-----IGAASSS 1301
Qy 1389 TLQGTLQNSLKFKLTEDELPGFIHGLTHGLTPAELLQKGIEHQMKQGSKLTFSV---D 1444

Db 1302 AC--SLNNS--SLNASQGMGSCSGSTFLPSPSSRYWRSSSHHQNQONNHQQOSQQLHGS 1356
Qy 1445 TSANLDLRA-----GINLNEDGSKPNGVTA--------RVSAG---- 1474

Db 1357 SSSNTCLMASPARPRSLSSNSSDSDVPGQNAGGSPSLDERLRNFEENYERWSGGSSREHI 1416

```
Qy    108 RDLLARDDGETQHEAA--AP-----DAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVP  160
              :::  ||  :                ||  | |                  ||||
Db    183 GDRSSTQNIHQNHQSARVAPPQSWYEAATAATTAQLKS------------SGGSG------  225

Qy    161 TQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSP------PHI--PGSHHEIKE    210
              |                    ::                    ||  ||
Db    226 -----NAG--------------ASAAVGFTMSSSPINHHPQHPHLQNPQHPHYTSS     263

Qy    211 EPVG---------STSKATTAHADRVEIAQEDDD------SEFQQLHQQRLAR-       248
          ||                :    :: :            : |  |:   |:
Db    264 PVVGAGSCPSAAQGQPQIQSQSQTTAVHRSVAYAGSAADDLLNTATSRNMLLHSSKLNKL   323

Qy    249 ----ERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQ--          294
              |                     |  ::      :     | :
Db    324 LKGAGATGSGGERSGSESPGRAGGATPLTTTSTITNNSFSSNSLNNTITTATPTMPFTAS   383

Qy    295 ---SMLKGSG--AGV-----------TPLAVTLDKGKLQLAPDNPPALNTLLKQT       333
              |||  |     |||           ||
Db    384 GAAGSVGLGSGAEAGVCSNSGTASGDILNVAAVLAAAVDNG---VPTHPIRTRHNLH--   437

Qy    334 LGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTG   393
          :  | :::  : :::     | :         |
Db    438 -GRSTTSSSRSHSRSPSSY----SSSHSSSSSHSSSHS-HASSPVQSSGNCAMA--      486

Qy    394 SVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDK   453
              :  ::|||  |                               :|:   :
Db    487 ---EGRSSR-TVNSVTVTSNSS--------------NPSGTA                    510

Qy    454 IHILHPELGVWQSADKDTHSQLSRQADGK-----LYALKDNRTL---QNLSDNKSSE     502
           : :       | ::    :|  :|  :        :|::  :|       | |::
Db    511 --VTVSSAGVGGGCGSSSSSSSSSSSSSCLTANPVVHSEDNRPLAIRVRNLPARSSDT    568

Qy    503 KLVDKIKSYSVDQRGQVA-----ILTDTPGRHKM------SIMPSLDASPESHISLSLHFA  552
           :  ::|  :  |            | |   :          :::     | :|| | |
Db    569 SLKDGL-FHEYKKHGKVTWVKVVGQNSERYALVCFKKPDDVEKALEVSHDKHFFGCKIEV   627

Qy    553 DAHQGL------LHGKS------ELEAQSVA--ISHGRLVVADSEGKL              586
           |:|        ||         :|  | |    ||  |
Db    628 EPYQGYGYDVEDNEFRPYEAELDEYHPKSTRTLFIGNLEKDITAGELRSH------FEAFGEI  682

Qy    587 FSAAIPKQG--------------DG---NELKM--KAMPQHA--LDEHF             614
                 ||               | |    | |    || |:    ||::
Db    683 IEIDIKKQGLNAYAFCQYSDIVSVVKAMRKMDGEHLGSNRIKLGFGKSMPTNCVWID---   739

Qy    615 GHDHQIS-GFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHH   673
          | |:|:  | |:|: || || :|    || || |::  |: |  |:  | ::|:|||||
Db    740 GVDEKVSESFLQSQFTRFGAVTKVSIDR------NRQLALVLYDQV--              779

Qy    674 TNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDC----KQLKKGLDGAAYL    727
                : :    |  :: || :|    :|:          :     ||:  | |||
Db    780 QNAQA-AVKDM--RGTI-LRRKKLQV-------DFASRECQDAFYDKQEKQQOQSS--    824

Qy    728 LKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPE-------PGDALQGLNKDDKA-    777
           | |   | ::| | | :  |:  :  |  :          || :|| |  | :
Db    825 -GSNPRFSRYESSASSLQSRSRASSFSRHQNNSNDDCSPINTPGGASSGISSASNLIN   881

Qy    778 QAMAV----IGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHK   833
          || |     ||:|   | :  | : |            :| |:   | ||
Db    882 QSTSININIGTNACSAM-----------PAPSLASAVVS----CNVNAS              916

Qy    834 QNLYALTHEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGS  893
           |:  :     |  |              :|: || |   |::|::
Db    917 GTVPAST--------SMPSGVSSSS---SSLPMSPAAL-----AQRHRMVRNARQTV      957

Qy    894 QHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRL------------MQG---VKGK      934
           | : |   : | |  |   |:  |    |               |    |  |
Db    958 DCDFNEVGRLRFRSSEE---VSGGAGNSTQFEDVRCDSPVTARQGSAVNCFTGPTAAVGE  1014

Qy    935 VIPGSGLITVKLSAQTGGMTGAEGRKVSSKKFSERIRAYAFNPTMSTPRPIKNAAYATQHGW  994
          :   :|:   :  |  | |                            | :|  :
Db   1015 SIDGTLNNNQITGGAEGFTGSGGSILSRRRCGK--------TPK--                1050
```

```
Qy    995 QGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDE  1054
          :::                  ::    :|               :|    ::  | :|
Db   1051 ---DLHPVHNQRIQLAEQVE------------ECPSSGDEG--VVSPRKRIKMD        1087

Qy   1055 LEQSATRSYTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNV----NRSGQDLSKSLQ   1109
          :|:|  |:    |:|      |  ::  |:|  |  |::         :|: |: ||
Db   1088 YHHHHHHS------NASGVESTGEHSSINKPSP--LLLSNCDVIHDPLNR------KSEI  1133

Qy   1110 QAVHATPPSAESKLQSMLGHFVSA--GVDMSHQKGEIPLGRQRDPNDKTALTKSRLILDT  1167
            | |  |    :: || |     :    |:|:     ||    :      |::| |
Db   1134 RRVSETPSGSPS--IKFPGHLPSAPQSLMLSCRRPSIDVGA------LSALSSSSAF--  1182

Qy   1168 VTIGELHELADKRAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKAL 1227
                : :  | ::    :|::              |::|                |||
Db   1183 -----RHGIVGASSMDQQHMMNASAAAKRRVTTTMQQPSSSS------------TTNSSS  1226

Qy   1228 EANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKIKNTLLSLDSGESMSFSRSY  1287
               | : | :   : |       | :        | |               || |
Db   1227 GSGLGGIJSSLTPA-DEYHHVSRGRGHQLHSHHSHEAS------------GGESADGSRP-  1273

Qy   1288 GGGVSTV---FVPTLSKKVPVP---------------VIPGAGITLDRAYNLFSFRSTSG  1328
          ||  |     |                         ||||          :: || |
Db   1274 --GTPLCDERPEVLPTEPRRLPPPRERVRERTRDVMWLPLPKFGVLF--FQQQQRSSSG   1328

Qy   1329 GLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSG  1388
          |                                  | :                  :|
Db   1329 G---------GG----AGNSYLQQQLGGGSTGGLGC-----------IGAASSS        1358

Qy   1389 TLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSV--D    1444
           | |:|:|||        | |      | | |  |  ||| | |  ||:|| ||:
Db   1359 AC--SLNNS--SLNASQGMGSCSGSTFLPSPSRYWRSSHHQNQONNHQOQSOLHGS     1413

Qy   1445 TSANLDLRA--------GINLNEDGSKPNGVTA----------------RVSAG--     1474
          |||                                              ||| |
Db   1414 SSSNTCLMASPARPRSLSSNSSDSDVPQQNAGGSPSLDERLRNFEENYERWSGGSSREHI  1473

Qy   1475 --------------LSASANLAAG--SRERSTTSGQFGS---TTSASNNRPTFLN--     1510
                        |||  ||     |:| ||  |   :   |||   |
Db   1474 SGHTPSSATPSWQLSMHMNLSTGLNSHQTSSASGNSNSSSGTVSSASNSRHKFLDIDEL   1533

Qy   1511 ----------------------GVGAGANL-TAALGVAHSSTHEGK                1533
                                || |  | : | || | ::|:  |
Db   1534 QPSDIVKSVLAKKSVFDDFQRLNKNQWYDPSSSDFALGSSSNIVTGSSLVANVSRHPGG  1593

Qy   1534 PVGIFPAFTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTK  1593
          | |   |   |::  |         :  :|    |         |:| |  : :  |
Db   1594 P--------CSGNTSPALP-------NLAAATKATPIIGNCSGGLGNSTGS-KSAGLLQ   1635

Qy   1594 MLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKLVIRQQAADSHSMELGS  1653
          :|: |                |     :   :  :|                | :| |
Db   1636 RLSSLSPMN--SPQASMSPYNSPSPSSVGG------VTACLGQLTKPAAPGTASAGL-  1685

Qy   1654 ASHSTTYNNLSRINNDGIVELLHKHFDAALP-ASSAKRLGEMNNDPALKDIIKQLQSTP  1712
           |  |  :    :|  | :  |  |  | |   |:|                | || |
Db   1686 -SGGTAASSSSPAANSGPTKGLQYPFPSHPPLPNTAAPPAVQPAPPLPEMGKQSRLTG  1744

Qy   1713 FSSA--SVSMELKDG----LREQTEK-AILDGKVGREEVGVLFQDRNNLRVK-SVSVS   1762
          |||    |:  |         |: |   | |  | |  :|::  |  | ||   |||:
Db   1745 QSSGNNLTKSLSVPDGPQSSPARVQLQKSASVPGSTNVGAPSSLSLDSTTASVETSASIS  1804

Qy   1763 QSVSKSEGFNTPALL---------LGTSNS-AAMSMERNIGTINFKYGQDQ          1803
           |||      |||            | |:  || |                 |
Db   1805 SSTSNGNSSLTSAAIHVQKPQQSTFVEEEHTKKSGTSTSQSSSSSSKKISSTHDKLHSKH  1864

Qy   1804 N  1804

Db   1865 N  1865
```

```
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   SPLIT ENDS LONG ISOFORM.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kuang B., Wu S., Shin Y.-A., Luo L., Kolodzie] P.;
RT   "Split ends encodes large nuclear proteins that regulate neuronal cell
RT   fate and axon extension in the Drosophila embryo.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF221715; AAF34661.1; -.
SQ   SEQUENCE 5554 AA;  599188 MW;  4037E27833D0C622 CRC64;

Query Match             2.5%;  Score 240.5;  DB 5;  Length 5554;
Best Local Similarity   19.5%;  Pred. No. 0.00024;
Matches  422;  Conservative  256;  Mismatches  744;  Indels  739;  Gaps  106;

Qy     6 LGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAAAS-----LAAEG----KNRGKMP   55
Db    76 LVTQYQANNSTSLAN---------SNSSPSSVSASVFATAGGSSERSRNRRDRPY     123

Qy    56 RIHQPSTAADGISAAHQQKKSF------SLRGCLGTKK--FSRSAPQGQPGTTHSKGATL  107
Db   124 RNGSASVQGGGINSSNTTTTAACTAGGSGSGAIGTGTGLVGSGPGGVP------QAL    176

Qy   108 RDLLARDDGETQHEAA--AP-----DAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVP  160
Db   177 GDRSSTQNIHQNHQSARVAPPQSWYEAATAATTAQLKS-----------SGGSG------  219

Qy   161 TQQKRHQLNFGQMRQTMLSKMAHPASANAGDRLQHSP------PHI--PGSHHEIKE    210
Db   220 ------NAG----------------ASAAVGFTMSSSPINHHPQHPHLQNPQHPHYTSS   257

Qy   211 EPVG-------STSKATTAHADRVEIAQEDDD------SEFQQLHQQRLAR-         248
Db   258 PVVGAGSCPSAAQGQPQIQSQSQTTAVHRSVAYAGSAADDLLNTATSRNMLLHSSKLNKL  317

Qy   249 ----ERENPPQQPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQ--         294
Db   318 LKGAGATGSGGERSGSESPGRAGAGATPLTTTSTITNNSFSSNSLNNTITTATPTMPTIAS 377

Qy   295 ---SMLKGSG--AGV---------TPLAVTLDKGKLQLAPDNPPALNTLLKQT        333
Db   378 GAAGSVGLGSGAEAGVCSNSGTASGDILNVAAVLAAAVDNG---VPTHPIRTRHNLH--   431

Qy   334 LGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTG  393
Db   432 -GRSTTSSSRSHSRSPSSY-----SSHHSSSSHSSSHS--HASSPVQSSGNCAMA----  480

Qy   394 SVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDK  453
Db   481 -----EGRSSR-TVNSVTVISNSS----------NPSGTA---------           504

Qy   454 IHILHPELGVWQSADKDTHSQLSRQADGK------LYALKDNRTL----ONLSDNKSSE  502
Db   505 --VTVSSAGVGGGCGSSSSSSSSSSSSSSSSCLTANPVHSEDNRPLAIRVRNLPARSSDT  562

Qy   503 KLVDKIKSYSVDQRGQVA-----ILTDTPGRKM------SIMPSLDASPESHISLSLHFA 552
Db   563 SLKDGL-FHEYKKHGKVTWKVVGQNSERYALVCFKKPDDVEKALEVSHDKHFFGCKIEV   621

Qy   553 DAHQGL---------LHGKS------ELEAQSVA---ISHGRLVVADSEGKL         586
Db   622 EPYQGYDVEDNEFRPYEAELDEYHPKSTRTLFIGNLEKDITAGELRSH------FEAFGEI 676

Qy   587 FSAAIPKQG-------------DG-----NELKM--KAMPQHA--LDEHF            614

Db   677 IEIDIKKQGLNAYAFCQYSDIVSVVKAMRKMDGEHLGSNRIKLGFGKSMPTNCVWID--  733
Qy   615 GHDHQIS-GFFHDDHGQLNALVKNNFRQQHACPLGNDHQPHPGWNLTDALVIDNQLGLHH  673

Db   734 GVDEKVSESFLQSQFTRFGAVTKVSIDR------------NRQLALVLYDQV--        773
Qy   674 TNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDC----KQLKKGLDGAAYL   727

Db   774 QNAQA-AVKDM--RGTI-LRRKKLQV-------DFASRECQDAFYDKQEKQQQQSS--   818
Qy   728 LKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPE------PGDALQGLNKDDKA--  777

Db   819 ---GSNPRFSRYESSASSLQSRSRASSFSRHQNNSNDDCSPINTPGGASSGISSASNLIN  875
Qy   778 QAMAV---IGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTLSREGISGELKDIHVDHK   833

Db   876 QSTSINISNIGTNACSAM------PAPSLASAVVS----CNVNAS                910
Qy   834 QNLYALTHEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGS  893

Db   911 GTVPAST---------SMPSGVSSSS----SSLPMSPAAL-----AQRHRMVRNARQTV  951
Qy   894 QHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRL-------MQG-----VKGK        934

Db   952 DCDFNEVGRLRFRSSEE---VSGGAGNSTQFEDVRCDSPVTARQGSAVNCFTGPTAAVGE 1008
Qy   935 VIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGW  994

Db  1009 SIDGTLNNNQITTGGAEGFTGSGGGSILSRRRCGK-----------TPK          1044
Qy   995 QGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDLQSKLETLDLGERGAELLNDMKRFRDE 1054

Db  1045 ----DLHPVHNQRIQLAEQVE-----ECPSSGDEG--VVSPRKRIKMD           1081
Qy  1055 LEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNV----NRSGGDLSKSLQ 1109

Db  1082 YHHHHHHS------NASGVESTGEHSSINKPSP--LLLSNCDVIHDPLNR----KSEI  1127
Qy  1110 QAVHATPSAESKLQSMLGHFVSA--GVDMSHQKGEIPLGRQRDPNDKTALTKSRLILDT  1167

Db  1128 RRVSETPSGSPS--IKFPGHLPSAPQSLMLSCRRPSIDVGA-----LSALSSSAF--  1176
Qy  1168 VTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNRAL 1227

Db  1177 -----RHGIVGASSMDQQHMNASAAAKRRRVTTMQQPSSSS---------        1220
Qy  1228 EANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSY 1287

Db  1221 GSGLGGISSLTPA-DEYHHVSRGRGHQLHSHSHEAS--------GGESADGSRP-    1267
Qy  1288 GGGVSTV----FVPTLSKKVPVP------VIPGAGITLDRAYNLSFSRTSG         1328

Db  1268 -GTPLCDERPEVLPTEPRRLPPPRERVRERTRDVVMLPLPKFGVLF-FQQQOSRSSG   1322
Qy  1329 GLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSG 1388

Db  1323 G---------GG-----AGNSYLQQQLGGGSTGGLGC---------IGAASSS       1352
Qy  1389 TLQGTLQNSLKFKLTEDELPGFIHGLTHGLTPAELLQKGIEHQMKQGSKLIFSV----D 1444

Db  1353 AC--SLNNS--SLNASQGMGSCSGSTFLPSPSSRYWRSSSHHQNQQNHHQQQSQQLHGS  1407
Qy  1445 TSANLDLRA-----------GINLNEDGSKPNGVTA------------RVSAG-----  1474

Db  1408 SSSNTCLMASPARPRSLSSNSSDSDVPGQNAGGSPSLDERLRNFEENYERWSGGSSREHI 1467
Qy  1475 ----------LSASANLAAG--SRERSTTSGQFGS-----TTSASNNRPTFLN-----  1510

Db  1468 SGHTPSSATPSWQLSMHMNLSTGLNSHQTSSASGNSNSSSGTVSSSASNSRHKFLDIDEL 1527
Qy  1511 -----------GVGAGANL-TRALGVAHSSTHEGK                          1533

Db  1528 QPSDIVKSVLAKKSVFDDDFQRLNKNQWYDPSSSDFALGSSSNIVTGSSLVANVSRHPGG 1587
```

Wed Jun 6 10:19:13 2001

us-09-596-784-2.rspt

```
Qy   1534 PVGIFPAFTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTK 1593
          |:    :|  | |         ::  |  :   |   :  : ||:||:|
Db   1588 P------CSGNTSPALP-------------NLAATKATPIGNCSGGLGNSTGS--KSAGLLQ 1629

Qy   1594 MLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSHSMELGS 1653
          :|:||          |    :          |   |:       :
Db   1630 RLSSLSPMN--SPQASMSPYNSPSPSVGG-----VTACLGQLTKPAAFGTASAGL-- 1679

Qy   1654 ASHSTTYNNLSRINNDGIVELLHKHFDAALP-ASSAKRLGEMMNNDPALKDIIKQLQSTP 1712
            ::|         |                :  | |   |:  :::  :|
Db   1680 -SGGTAASSSSPAANSGPTKGLQYPFPSHPPLPNTAAPPAVQPAPPPLPEMGKQSRLTG 1738

Qy   1713 FSSA--SVSMELKDG----LREQTEK-AILDGKVGREEVGVLFQDRNNLRVK-SVSVS 1762
          |::   :  :|  :|     ||  |  :|   | |||:: |:    |:  :| |:|
Db   1739 QSSGNNLTKSLSVPDGPQSSPARVQLQKSASVPGSTNVGAPSSLSLDSTTASVETSASIS 1798

Qy   1763 QSVSKSEGFNTPALL-------------------LGTSNS-AAMSMERNIGTINFKYGQDQ 1803
          | :          |                   |  |  :      |        :|  :|
Db   1799 SSTSNGNSSLTSAAIHVQKPQQSTFVEEEHTKKSGTSTSQSSSSSSKKISSTHDKLHSKH 1858

Qy   1804 N 1804
            |
Db   1859 N 1859


RESULT 14
Q92HL3

ID   Q92HL3;          PRELIMINARY;          PRT;     4152 AA.
AC   Q92HL3;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   LARGE SUPERNATANT PROTEIN 1.
GN   LSPA1.
OS   Haemophilus ducreyi.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=730;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=35000;
RX   MEDLINE=99030326; PubMed=9811662;
RA   Ward C.K., Lumbley S.R., Latimer J.L., Cope L.D., Hansen E.J.;
RT   "Haemophilus ducreyi secretes a filamentous hemagglutinin-like
RT   protein.";
RL   J. Bacteriol. 180:6013-6022(1998).
DR   EMBL; AF057695; AAC79757.1; -.
SQ   SEQUENCE 4152 AA;  456179 MW;  7D82DEDC988AB8F3 CRC64;

Query Match             2.5%;  Score  237.5;  DB 2;  Length 4152;
Best Local Similarity  18.9%;  Pred. No. 0.00022;
Matches  345;  Conservative  244;  Mismatches  636;  Indels  599;  Gaps  88;
```

```
Qy    350 GSQHLLLDNKGHLFDIKSTATSYSVLHNSHP--------------GEIKGKLAQAGTGSVSV 397
             : || : | |  ::  |  :| |  ::                 | | :  ||  |::|
Db    835 GGQGLNLTAKGNITN-DSNSTAIAVLHSNNDINLNANNKVYNIGEIYSQ------AGNISV 888

Qy    398 DGK--------SGKISLGSCTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRL 449
          || |        :||: | : |    |       ::: |
Db    889 EAKLLHNDVKLSGNIT--TTTKSGNATVKT---------------NSIGGLHDANSIRV 931

Qy    450 HDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIK 509
            |  |:| | | :| |      |:| ::|  |:|    : |   |       :::|| |
Db    932 GELTLN-------GKFADLD--NQLKVALRGKIYA-GSNLTFK---AKEGEKRQKSTA 976

Qy    510 SYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHI-SLSLHFADAHQGLLHGKSELEAQ 568
           |:   : |  :||  | |:| |   :|| :| :|   |  : |  | :  :  |
Db    977 QAKINRGTINV------KNKLEYGSNVDV--ENNMRSMQVNL--YEKIFNGDNPI--- 1022

Qy    569 SVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHF----GHDHOISGF 623
          || |: :|:| | :| :|| |  |  |||::|:|||   ||| ||       |  |:|||
Db   1023 TLTLKNGVTFAKDFSNRRRAS-NDGEGTNKKTFDNVAHLIEEAFSGYSNGNDHRAS-- 1078
```

```
Db    624 FHDDHGQLN---------ALVKNNFRQQHACPLGNDHQFHPGWN-LT----DALVID--NQ 668

Qy    669 LGL-----------HHTNPEPH---EILDMGHLGSLALQEG------------KL 697
          |:|                 :||   :|   |:| |  ::               |:
Db   1079 --DDGHVKSPYYLLVLAQAVNNTEGENYLKTALQHIFGPNWNDLTTTNNDTTINDKWNQ 1135

Qy    698 HYFDQLTKGWTGAESDCKQLKKG-LDGAAYLLKDGE----------------VKRLNINQS 741
          |  :: | |   :| :|  | :  :|  | ::|                    |||| | |
Db   1136 LKLKWEKFKNNGENNHSINLNIYPADEGVEKAKIFAGVLRNGTNGVEDKVYQELNDKAKK 1195

Qy    742 TSSIKHGT--ENVFSLPHVRNKPEPGDALQGLNK-------------------GIDKSIISELLAQPIYVAKADVPDVPRV 773
          | |:| |    | : | |  :|  |  |  |:|:                   || : | ||||| |  || |  | |  |
Db   1196 EYEDKFAKKFQGRFK--SRFQNGEFDWAGDWAKEGENESYGSKETEEKYNGIKKEHTVNIG 1253

Qy    774 --DDKA------QAMAVIGVNKYLAL--------ALTHE----------TEKGDIRSFQIKPGTQQLERPAQT 815
             ||        ::| |:|| |:            | |            |||| :  | |||   :|  | |
Db   1254 KHEIKVPTVSFENLNNINHQQDKSD------------GIDKSIISELLAQPIYVAKADVPDVPRV 1307

Qy    816 LSREGISGELKDIHVDHKQNLY-ALTHE-------------------GEVFHQPREAWQNGA 857
          :   | |:|  |: :| |::|  |  |                     |  |    | :|:
Db   1308 AQNDKAVDEDGLYRTRLSYINQNNYLGAKYFFNQLDTEDDKLKGIK-RIGDNYFEH--QL 1364

Qy    858 ESSSWH----------KLALPQSESKLKSLDMSHEHKPIATFEDGSQHLKAGGWHAYAAPE 909
           | ::           ::  ||     |:  :| ::  :   |          |:
Db   1365 ITR--LIEKVADNHLTLKHGLHDIALVKKLLIDSASIQAKDLNLKVGEAL--TKEQKDNLK 1420
```

```
Qy    910 RGP--------LAVGTSGSQTVFN------------RLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEG 957
          | |          | |: | ::|             ||:|| || :    |:| |:|| | |:
Db   1421 EDIWNYVKTEVNAQEVLVPQVYLAKOTIEEVEKQRGVGT----GQIRAGIIDVKVDDV 1474

Qy    958 RKVSSKFSERITRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAH 1017
          || ||   | |   |  |  :  |  :|   : | |      |  |:||:  | :||:|
Db   1475 RNTGTIAGYAVGLEAKNKLKNTGDILSQRLSKLVGKKGLESTGVT--YVDETG--ATKV 1529

Qy   1018 NVRHN-APQPDLQSKLETLDLGEHGAELLNDMKRFR----------DELEQSATRSVTV 1065
          :       | |  |:         |  :|    |             :| | | :| |
Db   1530 RKARIKSEGHI----------YLETDKD-KNVD-LTASELKGNTG--------QIKAK 1567

Qy   1066 LGQHQGVLKSNGEINSEFKPSPGKAL--------------VQSFNVNRSGQDLSKSL-QQAV 1112
          |  :|   |:  |::  | |  |:              |  :|:| |  | |    :||
Db   1568 DLNLNDIYETSYKYKYEKL-FGKNGGEIGDRVTQTSQAKSVGTDASFDHLHLSLEGDVNQ 1626

Qy   1113 HA-TPPSAESKLQSMLGHFV---SAGVDMSHOKGEIPLGRQRDPNDKTALTKSRLILDT 1167
          |   | :| | ::|  | |    : |:|||   | :  | |  | |:  || : |::
Db   1627 TGSNLKANRTTGVVKGDFNTKAGKDLFHRQIDTVTSGTVYSASASGGGQSAGISLTDGV 1686

Qy   1168 VTIGEL-----------HELADKAKLV------------SDHKPDA 1190
          ||:            |    | |             : |  |
Db   1687 ETYTNKTATAGANADVTNFMKRTRETETSLTHRNSEF------NALSGELYVMGK 1735

Qy   1191 DQIK--------QLRQQFDTLREKRYESNPVKHYTD-MGFT-------HNKA--LEA 1229
          | :        ::  | :  || || : || :| : ||    ||  |      | |
Db   1736 ADIGGVDINRDVEVIKTPEEIAAEQKAAEEAAKKAEVKENEASETAAKETEEAENDNVAEK 1795

Qy   1230 NYDAVKAFINAFKKEH----------HGVNLTRTVLESQGS------A 1262
          |    :: | | |:|           |  ||::| ||: | |
Db   1796 DKTKPFKFKLTDEIAAAFETKGEDFFAAYKAREEDRKKGFTLSAEQIESTKARDEKET 1855

Qy   1263 ELAKKLKNTLLSLD-SGESMSFSRSYGGG------------VSTVFVPTLSKKVVVP 1306
          |   |||:    | | |:|:|  :| |||            |:|  |||:  ::| |
Db   1856 TYYELKVGVGAEAEAHSAAADAISNKARQIIDTQNGLKQDGTVALQEASDVLNLATGDLA 1915

Qy   1307 VIPGAGIT--LDRAYNLSFSRTSG--GLNV---SFGRDGGVSGNIMVATGHDVMPYM 1356
          |:|:   |   | |  |  |: |    |:    : ||:|| |:| |  :| ||  |
Db   1916 GASAKLKFELSTIEKKSRGASDGRSILGGRLNLAARGGDITLNNVETTENSHLSLKARDN 1975

Qy   1357 TGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPFIHGLTH 1416
          | ||::||:| ||:||| ||:||  | | :||:|  :| :  | :| || | | ||  |
Db   1976 VNVNSGVTEQKDESNSQSLKVTAGASSGCGVMAGGCSAGVSAGVSGS--------YN 2024

Qy   1417 GHDHOISGF 1416
           |  |:|||
Db   2025 GKDFNLISSNVDVDHLHLDVKGDTN 2072
```

... too dense ...

```
Db  1337 ----------DDDEFARLASGSGTV----------PGAGATS--TCNKSGVGHTDSASA 1373
Qy  1237 FINAFKKEHHGVNLTT---RTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVST 1293
Db  1374 ------SRLHNPDVATAEDTEVIGTVGDVEQAEEVFRGQIEKD---YNFSTSV-DLSA 1421
Qy  1294 VFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVM 1353
Db  1422 LFAPSMGSGMP-PVAPPVA---------AASGLSI---GGASSS---SSGLALQ 1459
Qy  1354 PYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGT-LQGTLQNSLKFKLT------ED 1405
Db  1460 TATTRLHNSASLSSP------------RESAAASGSADLKGT-PRSLPPAMTSGLHTDED 1506
Qy  1406 ELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTSAN-LDLRA-GINLNEDG-S 1462
Db  1507 NIAAHAQPLPRSKAAPAD----------------ADMSANEAELRCRQVRLADGGAA 1547
Qy  1463 KPNGVTA----------------RVSAGLSASANL-----AAGSRERSTTSGQFGS 1497
Db  1548 EPEGRTASGGGDAAAAEAEAAHAGTETVAADEAATAQLSPVSPTAAASGERAATV-QTAA 1606
Qy  1498 TTSASNNRPTFLNGVGAGANLTAALGVAHSS---THEGKPVGIFPAFTSTNVSAALALDN 1554
Db  1607 APPASQDAVT-RNSIVLGATDGRGRGVGHLTYVPVHEG-PLFFVQSITATRVPVIKLTDK 1664
Qy  1555 RTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSAT--------TKMLAALKELDDAKP 1606
Db  1665 ATGTKVDITFAGGEHWRS--MQLTRSLLEVFPHARTLILFLKYCVRSLSAGESEPGGVT 1721
Qy  1607 AEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSHSME--------LGSASHS 1657
Db  1722 SFAIYLMVLHF----------YNECRQRILLLRERNAASSSAEAQGRQAGADAGEHQHQ 1771
Qy  1658 TTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMMNND------PALKDIIKQLQST 1711
Db  1772 ------------HQQLQKDRAAEEALRRGSAADNSAASPAAPRLAAVYQDASTT 1813
Qy  1712 ---PFSSASVSMELKDGLREQTEK-----------AILDGKVGREEVGVLFQDRNNL 1754
Db  1814 AGPAPLITGLLLEYMLGEYLARVEQRHAQLVEEAGARRSASLDA-AGAGDVAELYSSSGNI 1872
Qy  1755 RVKSVSVSQOSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYGQD 1802
Db  1873 Q-KTV----------PA--LGRASSSTAVATGACATIKAAYASD 1903
```

Search completed: June 5, 2001, 18:23:10
Job time: 391 sec

```
                    GenCore version 4.5
         Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     June  5, 2001, 18:17:04 ; Search time 28.9 Seconds
                                       (without alignments)
                                       2178.601 Million cell updates/sec

Title:           US-09-596-784-2
Perfect score:   9448
Sequence:        1 MELKSLGTEHKAAVHTAAHN........NPQVASALTDLKKEGLEMKS 1838

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters:    93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	232	2.5	1577	1	HLYA_PROMI	P16466 proteus mir
2	227.5	2.4	2541	1	TALI_HUMAN	Q9Y490 homo sapien
3	221	2.3	1608	1	HLYA_SERMA	P15320 serratia ma
4	219.5	2.3	2541	1	TALI_MOUSE	P26039 mus musculu
5	217.5	2.3	1957	1	YD86_SCHPO	Q10411 schizosacch
6	214	2.3	1781	1	AKAC_HUMAN	Q02952 homo sapien
7	207.5	2.2	3591	1	FHAB_BORPE	P12255 bordetella
8	204	2.2	2845	1	APC_MOUSE	Q61315 mus musculu
9	198.5	2.1	835	1	INVO_PONPY	P14708 pongo pygma
10	196.5	2.1	1902	1	P2P_LACPA	Q02470 lactobacill
11	195	2.1	3321	1	KEND_HUMAN	O95613 homo sapien
12	194	2.1	1861	1	MAP2_RAT	P15146 rattus norv
13	194	2.1	3358	1	PGCV_MOUSE	Q62059 mus musculu
14	193.5	2.0	1902	1	P3P_LACLC	P15292 lactococcus
15	189	2.0	1902	1	P2P_LACLA	P15293 lactococcus
16	189	2.0	2175	1	HMCU_DROME	P10180 drosophila
17	188.5	2.0	1857	1	MYST_HUMAN	P35749 homo sapien
18	188.5	2.0	2717	1	ZEP1_HUMAN	P15822 homo sapien
19	188	2.0	1230	1	ST20_CANAL	Q92212 candida alb
20	187	2.0	2022	1	ANT1_ONCVO	P21249 onchocerca
21	185.5	2.0	1325	1	G160_MOUSE	P55937 mus musculu
22	185.5	2.0	1935	1	MYSB_RAT	P02564 rattus norv
23	184.5	2.0	3210	1	CENF_HUMAN	P49454 homo sapien
24	183.5	1.9	2230	1	GOG4_HUMAN	Q13439 homo sapien
25	183	1.9	3866	1	HRX_MOUSE	P55200 mus musculu
26	182	1.9	1902	1	P1P_LACLC	P16271 lactococcus
27	181	1.9	1934	1	MYSB_MESAU	P13540 mesocricetu
28	179	1.9	2375	1	ATRX_HUMAN	P46100 homo sapien
29	178	1.9	1756	1	TRI2_ECOLI	P22706 escherichia
30	178	1.9	1959	1	MYSN_CHICK	P14105 gallus gall
31	178	1.9	1972	1	P531_HUMAN	Q12888 homo sapien
32	177.5	1.9	2326	1	PGG2_RAT	Q00657 rattus norv
33	176.5	1.9	918	1	YMJB_CAEEL	P34487 caenorhabdi
34	176	1.9	1966	1	MYSB_CAEEL	P02566 caenorhabdi
35	175	1.9	1935	1	MYSB_HUMAN	P12883 homo sapien
36	174.5	1.8	1972	1	MYST_RABIT	P35748 oryctolagus
37	174	1.8	2843	1	APC_HUMAN	P25054 homo sapien
38	173	1.8	2334	1	WAPA_BACSU	Q07833 bacillus su
39	173	1.8	2493	1	CYAA_USTMA	P49606 ustilago ma
40	172	1.8	2300	1	CYAA_NEUCR	Q01631 neurospora
41	171.5	1.8	2663	1	CENE_HUMAN	Q02224 homo sapien
42	171	1.8	1935	1	MYSB_PIG	P79293 sus scrofa
43	171	1.8	1939	1	MYSA_HUMAN	P13533 homo sapien
44	171	1.8	2842	1	APC_RAT	P70478 rattus norv
45	171	1.8	6359	1	BACC_BACLI	O68008 b bacitraci

ALIGNMENTS

```
RESULT 1
HLYA_PROMI
ID   HLYA_PROMI     STANDARD;      PRT;   1577 AA.
AC   P16466;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-NOV-1990 (Rel. 16, Last annotation update)
DE   HEMOLYSIN PRECURSOR.
GN   HPMA.
OS   Proteus mirabilis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Proteus.
OX   NCBI_TaxID=584;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.
RC   STRAIN=ISOLATE 477-12;
RX   MEDLINE=90170827; PubMed=2407716;
RA   Uphoff T.S., Welch R.A.;
RT   "Nucleotide sequencing of the Proteus mirabilis calcium-independent
RT   hemolysin genes (hpmA and hpmB) reveals sequence similarity with the
RT   Serratia marcescens hemolysin genes (shlA and shlB).";
RL   J. Bacteriol. 172:1206-1216(1990).
CC   -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC       CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC       DEFINED.
CC   -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC       ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA
CC       REQUIRES HPMB FUNCTION.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC   -!- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA
CC       MAY BE RESPONSIBLE FOR PORE FORMATION.
CC   -!- SIMILARITY: TO S.MARCESCENS HEMOLYSIN (SHLA).
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M30186; AAA25657.1; -.
DR   PIR; A35140; A35140.
KW   Hemolysis; Toxin; Outer membrane; Signal.
FT   SIGNAL        1     29
FT   CHAIN        30   1577       HEMOLYSIN.
SQ   SEQUENCE   1577 AA;  165869 MW;  175975E0C924B2D9 CRC64;

Query Match                      2.5%;      Score 232;   DB 1;   Length 1577;
Best Local Similarity   18.9%;  Pred. No. 0.0017;
Matches  329;  Conservative  238;  Mismatches  614;  Indels  564;  Gaps  84;

Qy   261 VATPISARFQPKLTAVAESVLEGTDTTQ------SPIKPQSMLKGSGAGVTPLAVTLDKG 314
             :||.|::  |::||  |::||:|:  ||||:  |:  :|  ::|:  :|
Db   197 IAPRIDSR-GKITAEISAFTGQNTFSQHFDILSSQKPVSALDSYFFG------SMQSG 248
```

```
Qy  1318 AYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHHIS 1377
                  :    :::   :       :    :::   : :  : :   :
Db  1028 A--LMETKAKGS---QFTSNGDISINVGENAHYEGAQFDAQKGKTVINAGG--         1073

Qy  1378 PDLRIGAAVS--GTLQGTLQNSLKFKL----TEDELPGFIHGLTH--------GTLT     1420
              :       :          ::    : : :  :   : :      :: :
Db  1074 -DLTLAQADTHSESQSNVNGSANLKVGTTPESKDYGGGFNAGTTHSKEQTAKVGTT     1132

Qy  1421 PAELLQKGIEHQMK-QGSKLTFSVDTSAN--LDLRAGINLNEDGSKPNGTARVSAGL     1475
              ::: :      :  :::::::       :  ::     :      :    : ::
Db  1133 GSQGIELNAGHNLTLQGTHLSSEQDIALNATNKVDLQSA-SSEHTEKGNNLSGGVQAGF   1190

Qy  1476 S---------ASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAAL     1522
         :         :   : : ::  ::   :  :     : :: :::: : :    :
Db  1191 GKKMTDDASSVNGLGSAQFAIGKQDEKSVSREGGTINNSGN-----              1232 L

Qy  1523 GVAHSSTH-EGKPVGIFPAFTSTNVSAALALDNR-TSQSISLELKRAEPVTSNDISELTS  1580
          : : ::  :::        : :::  : :  ::   :: :  :  :  : ::: ::
Db  1233 TINGNSVHLQGAQVN----------SKDTQLTSQSGDIETTSAQ---STDYKNNWG       1275

Qy  1581 T-LGKHFKDSATTKMLAALKELDDAKPAEQLHIL-------QQHFSAKDVVGDERYE      1629
         : ::  :   :    :   :  ::::    :  :       :::    :   : ::
Db  1276 TDIGFNGKKTNNTP-----KEVTEKPATSIHNIGGKLLVNVEDQQKTSHQNATLETGTL   1330

Qy  1630 AVRNLKKLVI-RQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPASS   1687
          :   :  :  ::  :   :  ::  ::   : ::  ::: :::   : :
Db  1331 TINSNKDLTLSGANVTADSVTGNVGGSLNIASQKESDRHVTVG-VNVGYNHTND-PKSS   1387

Qy  1688 ------AKRLGEMMNNDPALKDII---KQLQSTPFSSASVSMELKDGLREQTEKAIL     1735
               : : :   :       :   ::: : :   :   : ::: : :::: :  :  :
Db  1388 QVNKTAKAGGSLL-EKTIKDTIDSGIKSSTDAISDKYNSLSSTIADKTGISDET-KAKI  1444

Qy  1736 D---GKVG----------REEVGVLFQDRRNLRVKSVSVQSVSKSEGFNTPALL        1777
         :   ::::          :   :     : ::       ::::: :::: :::
Db  1445 DQGFGKVGNGIKNIVTGAEGHTANADIKVTHVD-NDAVTKTTSLTSNNDLSLNVNGSTKL 1503

Qy  1778 LGTS---------NSAAMSMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASALT   1827
         ::                :::    ::   :   :   :
Db  1504 TGAEIVSQOGQVDLGGSSVKLE-NIEGHHYEAGADLD                        1545

Qy  1828 DLKKE 1832
         :|::
Db  1546 DLAKQ 1550
```

```
RESULT 2
TALI_HUMAN
ID   TALI_HUMAN     STANDARD;      PRT;   2541 AA.
AC   Q9Y490; Q9UHH8;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TALIN.
GN   TLN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mao L., Fan Y.H.;
RT   "Complete cDNA sequence of human talin.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20079169; PubMed=10610730;
RA   Ben-Yosef T., Francomano C.A.;
RT   "Characterization of the human talin (TLN) gene: genomic structure,
RT   chromosomal localization, and expression pattern.";
RL   Genomics 62:316-319(1999).
CC   -!- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL
CC       STRUCTURES TO THE PLASMA MEMBRANE. HIGH MOLECULAR WEIGHT
CC       CYTOSKELETAL PROTEIN CONCENTRATED AT REGIONS OF CELL-SUBSTRATUM
```

```
Qy   315 KLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFLFDIKSTATSYSV 374
         :::     ::                 : :: ::::
Db   249 RIRI------INT------AEGSGVKLAGKFTADNDLSVKADN--------IQTDSQV      286

Qy   375 LHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLSGSTQSHNKTMLSQPGEAHRSLLTGIWQ   434
         :::    :    :    :  :  :::    :: : :
Db   287 RYDSYDKD--------GSENYQNYRGGITVNNSGSSQTLTKTELKGK-NITLVASSH      334

Qy   435 HPAGAARPQGESIRLHDDKIHILHPELG------------VWQ---SADKDTHSQLSRQA   479
         :::  : ::  : :     :     : :             : :   : : :::  ::
Db   335 NQIKASDLMGDDITLQGADLTIDGKQLQQKETDIDNRWFYSWKYDVTKEKEQIQQIGSQI   394

Qy   480 DGK---LYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMP   535
          :    :  : : : :    :   :  :  :             : :: :
Db   395 DAKNNATLTATKGDVTLDAAKINAGNNLAINANKDIHING-----LVEKESR-------  441

Qy   536 SLDASPESHISL-SLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPK   593
         : :::   : :   :     :  :   :  ::    : :::   :  :::  :  : :
Db   442 SENGNKRNHTSRLESGSWSNSHQTETLKASELTA-------GKDLGLDAQGSITAGGAKL  494

Qy   594 QGDGNELKMKAMPQHALD-EHFGHDHQISGFFHDDH---------GQLNALVKNNFRQQ--- 642
          :    :::     : :    : :::  ::  :            :: : :  ::
Db   495 HANENVL-VNAKDNINLNVQKTNNDKTVT----DNHVMWGGIGGQN--KNNNNQQQVS     546

Qy   643 HACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQ   702
         :          :       :                      : :
Db   547 HATQLTADGQL-----------LLAADNNVNI----------------------       567

Qy   703 LTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLN-INQSTSSIKHGTENVFSLPHVRNK  761
         : ::     ::     :      : : ::            : : : :
Db   568 ----TGSQ----VKGNQGAFVKTTQGDVVIDNALSETISKIDERTGTAFNITKSSHK     616

Qy   762 PEPG-DALQGLNKDDKAQAMAVIG-VNKYLALTEKGD------IRSFQIKPGTQ--QLE   810
         : :   : :::  ::  :          : :   ::        : : ::::::   ::
Db   617 NETNKQTSGSELISDAQLTVVSGNDVNVIGSLIKSADKLGIHSLGDINVKSAQQVTKID   676

Qy   811 RPAQTLSREGISGELKD--------IHVDHKQNLYALTHEGEVFHQPREAWQNGAESSSWHK 864
         ::   : :::  ::::        :  : ::   :::  :  :::
Db   677 DEKTSLAITGHAKEVEDKQYSAGFHITHTTNKNTSTE-----------------       713

Qy   865 LALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSGTVF  924
         : :     :  ::  : :       :   :                 : : :: ::: :
Db   714 --TEQANSTISGANVDLQANKDVTFA-GSDLKTTAGN----------ASITGDNVAF    757

Qy   925 ----NRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTP  980
              : : :::  ::::     :   :                 :  :     : :
Db   758 VSTENKKQTDNTDTTISGG------FSYTGGVD-----KVGSKAD----FQYDKQHTQT  801

Qy   981 RPIKNAAYATQ---------HGWQQREGLKPLYEMQGALIKQLDAHNVRH            1021
         ::::                 ::       :               :  :
Db   802 EVTKNRGSQTEVAGDLITTANKDLLHEGASHHVEGR------YQESG------ENIQH   847

Qy  1022 NAPQPDLQSKLETLDLG-EHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN  1080
         :  :   :: :: ::::     :::                            :  :
Db   848 LAVNDSETSKTDSLNVGIDVGVNL-------------DYSGVTK-------          878

Qy  1081 SEFKPSPGKALVQSFNVNRSGQ-DLSKSLQQAVHATPSAESKLQSMLG-HFVSAGVDM   1137
               : ::      : ::::   : : : ::          ::  :      : :::
Db   879 -----PVKKAIEDGVNTKPGNNTDLTKKV-------TARDAIANLANLSNLETPNVGVEV 927

Qy  1138 SHQKGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLR  1197
         :     :::     :::   : : ::: :      :    : :: :::  :: ::::::
Db   928 GIKGG-----GSQQSQTDSQAVSTS--------INAGKIDIDSNNKLH            962

Qy  1198 QQFDTLREKRYESNPVKHYHYTDMGFTHNKALEANVDAVKAFINAFKKEHHGVNLTRTVLE 1257
         ::                :: :::         ::                :: :::::
Db   963 DQ--------GTH------YQSTQEGISLFANT---                         981

Qy  1258 SQGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDR  1317
         : :::::::::::     : : : ::   ::  :::      ::::  : : ::  :: ::
Db   982 -------HTSEATLDKHQT-TFHETKGGG--QIGVSTKTGSDITVAIKGEGQTTDN    1027
```

```
CC    CONTACT AND, IN LYMPHOCYTES, AT CELL-CELL CONTACTS (BY
CC    SIMILARITY).
CC -!- SUBUNIT: BINDS WITH HIGH AFFINITY TO VINCULIN AND WITH LOW
CC    AFFINITY TO INTEGRINS (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,
CC    RADIXIN, AND TALIN.
CC -!- SIMILARITY: IN THE C-TERMINUS TO YEAST SLA2 AND C.ELEGANS ZK370.3.
CC
```
```
DR    EMBL; AF078828; AAD13152.1; -.
DR    EMBL; AF177198; AAF23322.1; -.
DR    MIM; 186745; -.
DR    InterPro; IPR000159; -.
DR    InterPro; IPR000299; -.
DR    InterPro; IPR002558; -.
DR    Pfam; PF00373; I_LWEQ; 1.
DR    Pfam; PF01608; Band_41; 1.
DR    PROSITE; PS00660; BAND_41_1; 1.
DR    PROSITE; PS00661; BAND_41_2; 1.
DR    PROSITE; PS50057; BAND_41_3; 1.
KW    Structural protein; Cytoskeleton.
FT    DOMAIN    173    332    BAND 4.1-LIKE.
FT    CONFLICT 1227   1227    L -> S (IN REF. 2).
FT    CONFLICT 1966   1966    R -> A (IN REF. 2).
SQ    SEQUENCE 2541 AA;  269927 MW;  0FEDA1A9E16E780 CRC64;

Query Match            2.4%; Score 227.5; DB 1; Length 2541;
Best Local Similarity 19.5%; Pred. No. 0.0058;
Matches 449; Conservative 297; Mismatches 846; Indels 707; Gaps 114;

Qy    11 KAAVHTAAHNPVG---HG-VALQ-QGSSSSSPQN------AAASLAAEGKNRGKMPR 56
Db   428 KSTVLQQQYNRVGKVEHGSVALPAIMRSGASGPENFQVGSMPPAQQOITSGQMHRGHMPP 487

Qy    57 IHQPSTADG-------ISAAHQQKKSFSLRGCLG----TKKFSRSAPGQGPGTTHSK- 103
Db   488 LTSAQQALTGTINSSMQAVQAAQATLDDFDTLPPLGQDAASKAWRKNKMDESKHEIHSQV 547

Qy   104 -----GATLRDLLARDDGETQHE-----AAAPDAARLTR-SGGVK--RRNMDDM-AGRP 148
Db   548 DAITAGTASVVNLTAGDPAETDYTAVGCAVTTISSNLTEMSRGVKLLAALLEDEGGSGRP 607

Qy   149 MVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRIQH-----SPPHIPG 203
Db   608 LLQAAKGLAGAVSELLRSAQPASAEPRQNLLQAAGNVGQA-SGELLQQIGESDTDPHFQD 666

Qy   204 SHHEIKEEPVGSTSKATAHADRVEIAQEDDSEFQQLHQQRLARERENPPQPPKLGVAT 263
Db   667 ALMQL-AKAVASAAAALVLKAK--SVAQRTEDSGLQ---TQVIAAATQCALSTSQLVACT 720

Qy   264 PISARFQPKLTA------------VAESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAVT 310
Db   721 KVVA--PTISSPVCQEQLVEAGRLVAKAVEGCVSASQAATEDGQLLRGVGAAAT--AVT 775

Qy   311 LDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTAT 370
Db   776 -------QALNELL-----QHVKAHATGAGPAGR--------YD-QATDT 804

Qy   371 SYSVLHN----SHPGEIKGK---LAQAGT--GSVSVDGKSGKISLGSGTQSHNKTMLS 419
Db   805 ILTVTENIFSSMGDAGEMVGQARILAQATSDLVNAIKADAE-----GESDLENSRKLLS 858

Qy   420 QP--GEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLS 476
Db   859 AAKILADATAKMV---EAAKGAA------AHPD------SEEQQQRLR 891

Qy   477 RQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPS 536
Db   892 EAAEG--LRMATNAAAQ----NAIKKLLVQRLE-HAAKQAAASATQTIAAAQHAAS-TPK 943

Qy   537 LDASPESHISLSLH-FADAHQGLLHGKSELEAQSVAISHGRLVVADSE----GKLFSA 589
Db   944 ASAGPQPLLVQSCKAVAEQIPLLVQGVRGSQAOPDSPSAQLALIAASQSFLQPGKMVAA 1003

Qy   590 A---TPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN---- 637
Db  1004 AKASVPTIQD---QASAM--------QLSQCAKNLGTALAELR 1035

Qy   638 --NFRQQHAC-PLGNDHQFHPGWNL------TDALVIDNQLGLHHTNPEPHEIL-----D 683
Db  1036 TAAQKAQEACGPLEMDSALSVVQNLEKDLQEVKAAARDGKL-----KPLPGETMEKCTQD 1090

Qy   684 MGH------LGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDG 731
Db  1091 LGNSTKAVSSAIAQLLGEVA--QGNENYAGIAARDVAGG---LRSLAQAARGVAALTSDP 1145

Qy   732 EVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDKAQAMAVIGVNKYL- 789
Db  1146 AVQAIVLDTASDVLDKASSLIEEAKKAAGHPGDPESQORLAQVAKAVTQA---LNRCVSC 1202

Qy   790 -------ALTEKGD----IRSFQIKPGTQQLERPAQTLSREGISG----ELKDIHV 830
Db  1203 LPGQRDVDNALRAVGDASKRLLSDLLPPSTGTFQE-AQSRLNEAAAGLNQAATELVQASR 1261

Qy   831 DHKQNL-YALTHEGEVFHQPREAWQNGAESSSWHKLALPQSE---SKLKSLDMSHEH 883
Db  1262 GTPQDLARASGRFGQDFSTFLEA--GVEMAG---QAPSQEDRAQVVSNLKGISMSSSK 1314

Qy   884 ----KPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTV 923
Db  1315 LLLAAKALST-DPRAPNLKS----QLAAAAR--AVTDSINQLITMCTQQAPGQKECDN 1365

Qy   924 --------FNRLMQGVKGKVIPGSGLT-VKLSAQTG---GMT 953
Db  1366 ALRELETVRELLENPVQPINDMSYFGCLDSVMENSKVLGEAMTGISQNAKNGNLPEFGDA 1425

Qy   954 GAEGRKVSSKFPSER1RAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQ 1013
Db  1426 ISTASKALCGFTEAAAQAAYLVGVSDP--------NSQAGQQGLVEPTQFARA--- 1470

Qy  1014 LDAHNVRHNAPQPDLQSKLETLDLGEHG---AELLNDMKRFRDELEQSATRSVTVLGQHQ 1070
Db  1471 -------NQAIQMACQS----LGEPGCTQAVLS-------AATIVAKHT 1502

Qy  1071 GVLKSNGEINS-EFKPSPGKALVQSF-NVNRSGQDLSKSLQ------QAVHAT 1115
Db  1503 SALCNSCRLASARTTNPTAKRQFVQSAKEVANSTANLVKTIKALDGPFTEENRAQCRAAT 1562

Qy  1116 PPSAE-----------SKLQSMLGH------FVSAGVDMSHQGEIPLGR 1148
Db  1563 APLLEAVDNLSAFASNPEFSSIPAQISPEGRAAMEPIVISAQTMLESAGGLIQTARALAV 1622

Qy  1149 -QRDPNDKTALT-KSRLILDTV------TIGELH-ELA------DKAKLVS 1184
Db  1623 NPRDPPSWSVLAGHSRTVSDSIKKLITSMRDKAPGQLECETAIAALNSCLRDLDQASLAA 1682

Qy  1185 --------------------DH----KPDADQ----IKQLRQQFD-T 1202
Db  1683 VSQQLAPREGISQEALHTEMLTAVQEISHLIEPLAHAARAEASQLGHKVSQMAQYFEPLT 1742

Qy  1203 LREKRYESNPVKHYTDMGFTH------NKALEANYDAVKAFINAFKKEHHGVNLTTRTVL 1256
Db  1743 LAAVGAASKTLSHPQQMALLDQTKTLAESALQLLYTAKEAGGNPKQAAH-----TQEAL 1796

Qy  1257 ES--QGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVIPGAGIT 1314
Db  1797 EEAVQMMTEAVEDLTTTL------NEAASAAGVVGGMVDSI-TQAINQLDEGPMGEPEGSF 1850
```

```
Qy 1315 LDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTG-------------- 1358
         :|: |  |::  ||   ::     |
Db 1851 VD--YQTTMVRTAKAIAVTVQE------MVTKSNTSPEELGPLANQLTSDYGRLASE 1899

Qy 1359 ---KKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFLTEDELPGFIHGLT---------- 1415
         :|| :||   :: ||    |   | :|
Db 1900 AKPAAVAAENEEIGSHIKHRVQ-ELGHGCAALVTKAGALQCSPSDAYTKKEL-------- 1950

Qy 1416 HGTLTPAELLQKGIEH--QMKQGSKLIFSVDTSAN------LDLRAGI----NLNED 1460
         |  | || |    :    | | :|  |     :| ||
Db 1951 ---IECARRVSEKVSHVLRALQAGNRGTQACITAASAVSGIIADLDTTIMFATAGTLNRE 2007

Qy 1461 GSKP-----NGV-TARVSA-GLSASANLAAGSRERSTTSGQ------------FGST 1498
         |::         |   |||||:|                              |:
Db 2008 GTETFADHREGILKTAKVLVEDTKVLVQNAAGSQEKLAQAAQSSVATITRLADVVKILGAA 2067

Qy 1499 TSASNNRPTFLNGVGAGANLTAALGVAHSSTH-EGKPVGIFPAFTSTNVSAALALDNRT 1556
         |:::    : |||  ||:|          ||    | :| ||  |: :||  :: :|
Db 2068 SLGAEDPETQVVLINAVKDVAKALGDLISATKAAAGK-VGDDPAVWQLKNSAKVMVTNVT 2126

Qy 1557 S--------------------QSISLEL-------------KRAEPVTSNDISELTSTLGKHF 1586
         |                    :|:|:|              |:||  ||:|||:|: ||||
Db 2127 SLLKTVKAVEDEATKGTRALEATTEHIRQELAVFCSPEPPAKTSTPEDFIRMTKGI---- 2182

Qy 1587 KDSATTKMLA--------ALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLK 1635
         ||| ||               ||| : :: :|    :: |:||   ||:|||:||:
Db 2183 -TMATAKAVAAGNSCRQEDVIATANLSRRAIADMLRACKEAAYHPEVAPDVRLRALHYGR 2241

Qy 1636 KLV------IRQQAADSHSMELGSASH----------------------S 1657
         |:|      ||| | :|: ::|::||                        S
Db 2242 ECANGYLELLDHVLLTLQKPSPELKQQLTGHSKRVAGSVTELIQAEAMKGTEWVDPEDP 2301

Qy 1658 TYYNNLSRINNDGIVELLHKFEDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSAS 1717
         | |   |:| ||| | :  |:| |::   :||  || :::| ||  |:
Db 2302 TVIAENELLGAAAIEAAAKKLEQLKPRAKPKEADESLNFEEQILEAAKSI-----AAA 2355

Qy 1718 VSMELKDGLREQTEKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQSVSKS--------- 1768
         ||:::|  |:|: | || ||||||  ::||  |:::::| :: |:||||
Db 2356 TSALVKAASAAQRE-LVAQGKVGAIPANAL--DDGQWSQGLISAARMVAAATNNLCEAAN 2412

Qy 1769 ---EGFNTPALLGTSNSAAMSMERNIGTINFKYGQDQNTPRRFTLEG-GIAQAN------ 1819
            ||:|| ||| :::| :  |:|  |  :| |  |:|  :| :|||:|  :|:
Db 2413 AAVQGHASQEKLISSAKQVAASTAQLLVACKVKADQDSEAMKRLQAAGNAVKRASDNLVK 2472

Qy 1820 -PQVASALTDLKKEGLEMK 1837
          ||:| |||:|||:|| :|
Db 2473 AAQKAAAAFEEQENETVVVK 2491
```

RESULT 3

HLYA_SERMA

```
ID   HLYA_SERMA     STANDARD;      PRT;   1608 AA.
AC   P15320;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-NOV-1990 (Rel. 16, Last annotation update)
DE   HEMOLYSIN PRECURSOR.
GN   SHLA.
OS   Serratia marcescens.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Serratia.
OX   NCBI_TaxID=615;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
RC   STRAIN=SN8;
RX   MEDLINE=88257037; PubMed=3290200;
RA   Poole K., Schiebel E., Braun V.;
RT   "Molecular characterization of the hemolysin determinant of Serratia
RT   marcescens.";
RL   J. Bacteriol. 170:3177-3188(1988).
CC   -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
CC       CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
CC       DEFINED.
```

```
CC   -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC       ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
CC       REQUIRES SHLB FUNCTION.
CC   -!- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC   -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; M22618; AAA50323.1; -.
DR   PIR; A28182; A28182.
KW   Hemolysis; Toxin; Outer membrane; Signal.
FT   SIGNAL        1     30
FT   CHAIN        31   1608       HEMOLYSIN
SQ   SEQUENCE   1608 AA;  165078 MW;  D669B476FE7DAD51 CRC64;
```

```
Query Match
Best Local Similarity   2.3%;      Score 221;  DB 1;   Length 1608;
Matches  372;  Conservative  235;  Mismatches  689;  Indels  700;  Gaps  89;
           18.6%;           Pred. %:  0.0059;
```

```
Qy   33 SSSSPQNAAASLAAEGKNRGKMPRIHQPSTAA----------DGISAAHQQKKSFSLR--GC 82
Db   23 AASAGAYAAEIVAANGANG--PGVSTAATGAQVDIVAPNGNGLSHNQYQDFNVNQPGA 79

Qy   83 L--GTKKFSRSRAPQGQPGTTHSKG------ATLRDLLARDDG----------ETQHEAAA 124
Db   80 VLNNSREAGLSQLAGQLGANPNLGGREASVLNEVGRNPSLLHGGQEIFGMAADYVLAN 139

Qy  125 PDAARLITRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAH 184
Db  140 PNGISCQSCGFINTSHSSLVVGNPLVENGVLQG----------YSTFGNRNTLSLN---- 185

Qy  185 PASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVE-IAQEDDSEFQQLHQ 243
Db  186 -GTLNAGGVLDLLIAPKIDSRGEVIVQDFKQSNGKVTSAAINAISGLNRVARDGTVQASQO 244

Qy  244 QRLARERE------NPPQPPKLGVATPISARFQPKLTAV---------AESV 280
Db  245 MPTALDSYYLGSMQAGRINIINTAQGSGVKLAGSLNAGDELKVKAYDIRSESRVDDASSN 304

Qy  281 LEGTDTTQ----SPLKPQSMLKG----------SGAGVTPLAVTLDKG 314
Db  305 KNGGDNYQNYRGGIYVNDRSSSQTLTRTELKGKNISLVADNHAHLFATDIRGEDITLQGG 364

Qy  315 KLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSV 374
Db  365 KLTL----------------DGQQ--LKQTQGHTDDRWFYSWQYDV 392

Qy  375 LHNSHPGEIKGKLAQAGTGSVSVDGKSGKIS----------LGSGTQSHNKTMLSQPGEAH-- 425
Db  393 TRE------REQLQQAGS-TVAASGSAKLISTQEDVKLLGANVSADRALSVKAARDVHLA 445

Qy  426 RSLLTGIWQH--------PAGAARPQGE------SIRLHDDK 453
Db  446 GLVEKDKSSERGYQRNHTSLRTGRWSNSDESESLKASELRSEGELITLKAGRNVSTQGAK 505

Qy  454 IHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQ--NLSDNKSSEKLVDKIKS 510
Db  506 VHAQRD----LTIDADNQIQVGVQKTANAKAVRDDKTSWGGIGGGDNKNNS-- 552

Qy  511 YSVDQRGQVAILTDPGRHKMSIMPSLDASPESHISLSLHFADAHQLLHGKSELEAQ-- 568
Db  553 ------NRREISHASELTSGGTLRL------NGQQGVTITGSKARGQKG 589

Qy  569 -SVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDD 627
Db  590 GEVTATHGGLRIDNA---LSTTVDK----------IDARTGTAFNITSSSHK- 628
```

```
Qy  1569  -------PVTSNDIEL---TSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQH  1616
                 :::  :  |||          ::||| ::
Db  1452  GKVTLPATPAGEKPQGPLWDRGARTVGGAVKDSIT----------GPAGR---QGH  1494

Qy  1617  FSA------KDVGDERYEAVRNLKKLVIRQQA--------ADSHSMELGSASHSTTYNN  1662
                        :  |||                    :||:  :
Db  1495  LKVNADVVNNNAVGEQSAIAGKNGVALQVGGQTOLTGGEIRSQQGKVELGGSQVSQQDVN  1554

Qy  1663  LSRINNDGIVELLHKFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMEL  1722
                :  :|             ::||          |||
Db  1555  GQRYQGGGRVD----------------AAATVGLLGG--AAKQSV--AGNVPFASGHASTQQ  1597

Qy  1723  KDGLREQTEKAILDGK  1738
              :  |||
Db  1598  AD-------AKAGVFSGK  1608


RESULT  4
TALI_MOUSE

ID   TALI_MOUSE           STANDARD;         PRT;    2541 AA.
AC   P26039;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TALIN.
GN   TLN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Fibroblast;
RX   MEDLINE=91015390; PubMed=2120593;
RA   Rees D.J.G., Ades S.A., Singer S.J., Hynes R.O.;
RT   "Sequence and domain structure of talin.";
RL   Nature 347:685-689(1990).
CC   -!- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL
CC       STRUCTURES TO THE PLASMA MEMBRANE. HIGH MOLECULAR WEIGHT
CC       CYTOSKELETAL PROTEIN CONCENTRATED AT REGIONS OF CELL-SUBSTRATUM
CC       CONTACT AND, IN LYMPHOCYTES, AT CELL-CELL CONTACTS.
CC   -!- SUBUNIT: BINDS WITH HIGH AFFINITY TO VINCULIN AND WITH LOW
CC       AFFINITY TO INTEGRINS.
CC   -!- PTM: PHOSPHORYLATED.
CC   -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,
CC       RADIXIN, AND TALIN.
CC   -!- SIMILARITY: IN THE C-TERMINUS TO YEAST SLA2 AND C.ELEGANS ZK370.3.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X56123; CAA39588.1; -.
DR   PIR; S11661; S11661.
DR   HSSP; P04002; 1WFA.
DR   MGD; MGI:1099832; Tln.
DR   InterPro; IPR000299; -.
DR   InterPro; IPR002558; -.
DR   Pfam; PF00373; Band_41; 1.
DR   Pfam; PF01608; I_LWEQ; 1.
DR   PROSITE; PS00660; BAND_41_1; 1.
DR   PROSITE; PS00661; BAND_41_2; 1.
DR   PROSITE; PS50057; BAND_41_3; 1.
KW   Structural protein; Cytoskeleton; Phosphorylation.
FT   DOMAIN      173    332       BAND 4.1-LIKE.
FT   VARIANT    1105   1105       L -> P.
FT   VARIANT    2180   2180       K -> M.
SQ   SEQUENCE   2541 AA;  269832 MW;  14EF75ABE9FC2CBB CRC64;
```

```
Qy  628   HGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT-----DALVIDNQLGLHHTNPEPHEIL  682
              |||            :|| ||  ||  :
Db  629   ------ADNSYQSSTASELKSD------TNLTLVSHKDADVIGSQVA-------------  663

Qy  683   DMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQST  742
                     || :|| |   ||  :||
Db  664   -------------SGGE----------------------LSVESKT  674

Qy  743   SSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGD-------  796
             :|    ||    ||  ::| ::|    | ||  ||
Db  675   GNI-----NVKA----AERQQNIDEQKTALTVNGYAK-----EAGDKQYRAG  712

Qy  797   --IRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFHQPREAWQ  854
               :|| :|    ||  :  :||           ||  :
Db  713   LRIEHTRDSEKTTRTENSASSLS--GGSVKLK---------AEKDVTFSGSKLVAD  757

Qy  855   NGAESSSWHIKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLA  914
               :|: |||  ||   :           |  ||
Db  758   KGDASVSGNKVSFLAADDKTAS--------------NTEQTKIGGGFYY----------  792

Qy  915   VGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRA--Y  971
               |||   |      || |:|   ||    :||
Db  793   --TGG-----IDKLGSGVE-----------------AGYENNKTQAQSSKAITSGSDV  826

Qy  972   AFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDLQSK  1031
             :|  ::| | |             :|   ||        ||
Db  827   KGNLTINARDKL-----TQQGAQ-----HSVGGAY--QENAAGVDHLAAADTASTT  870

Qy  1032  LETLDLGEH-GAELLNDMKRFRDELEQSATRSVTVLGQHGGVLKSNGEINSEFKPSPGKA  1090
              :|: |||  :  || :   :|||     |  ::|   ||
Db  871   TTKTDVGVNIGANV--DYSAVTRPVERAVGKAAKL--DATGVINDIGGIGA---PNVG-  921

Qy  1091  LVQSFNVNRSGQDLSK--SLQQAVHATPPSAESKIQSMLGHFVSAGVDMSHQKGEIPLGR  1148
               |:  :||| |||  |   :|| |  :| :|  |    ||
Db  922   ---LDIGAQGGSSEKRSSSSQAVVSS-----------------VQAGSIDINAKGEV---  958

Qy  1149  QRDPNDKTALTKSRLILDTVTIGEELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRY  1208
                :| |||| |              ||  :|
Db  959   -RDQGTQYQASK----------GAVNLTAD------SHRSEAAANRQDEQSRDT-----R  996

Qy  1209  ESNPVKHYTDMG-------------------FTHNKALEANYDAVKAFINAFKKE--HHGVN  1249
              :| : ||         ::|::      || | ::    ||
Db  997   GSAGVRVTTTTGSDLTVDAKGEGGTQRSNSSASQAVTGSIDAANGINVNVKKDAIYQGTA  1056

Qy  1250  LT---TRTVLESQGSAEL-----AKKLKNTLLSLDSG--ESMSFSRSYGGGV-----  1291
             :|   | ||       :|     |  |||    |   :    :|  ||:
Db  1057  LNGRGKTAVNAGGDIRLDQASDKQSESRSGFNVKASAKGGFTADSKNFGAGFGGGTHNG  1116

Qy  1292  ---STVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFG----------  1335
                :|  || |   |  ||  |||      |
Db  1117  ESSSSTAQVGNISGQQGVELKAGRDLTLQG----TDVKSQGDVSLSAGNKVALQAAESTQ  1172

Qy  1336  --RDGGVSGNIMVATGHDVMPYMTGKKTSAGNASD-------------------WLS  1371
              ::  :|||:  ||   |   |        ||
Db  1173  TRKESKLSGNIDLGAGSSDSKEKTGGNLSAGGAFDIAKVNESATERQGATIASDGKVTLS  1232

Qy  1372  AKHKISPDLRI-GAAVSG---TLQ-----GTLQNSLKFKLTEDELPGFIHGLTHGTLTPAE  1423
             :|   : ||    |      :||     :|| |||
Db  1233  ANGKGDDALHLQGAKVSGGSAALEAKNGGILLESAKNEQHKDN---------------  1275

Qy  1424  LLQKGIEHQMKQGSKLTFSVDTSANLD----------LRAGINLNEDGSKPN-GVTA  1469
             ||   |||  :|||          : |          ||   :| ::|     :||
Db  1276  -WSLGIKANAKGGQ--TFNKDAGGKVDPNTGKDTHTLGAGLKVGVEQQDKTTHANTGITA  1332

Qy  1470  -RVSAGLSASANLAAGSRERSTTSGQFGSTTSASNNRPPFLNGV-----------GAGA  1516
              :|:  |:|:|| |        |||      | :| || |||          ||
Db  1333  GDVTLNSGKDTRLAGARVDADSVQGKVGGDLHV-ESRKDVENGVKVDVDAGLSHSNDPGS  1391

Qy  1517  NLTAALGVAHSSTHEGK------PVG---IFPAFTSTNVSAALALDNRTSQSISLELKRAE  1568
             :|:  ||| |||||      |||   :: :|||      | :||
Db  1392  SITSKLSKVGTPRYAGKVKEKLEAGVNKVADATTDKYNSVARRLDPQQDTTGAVSFSKAE  1451
```

```
Query Match     2.38;    Score 219.5;  DB 1;  Length 2541;
Best Local Similarity   19.4%;  Pred. No. 0.014;
Matches  444;  Conservative 295;  Mismatches 860;  Indels 693;  Gaps 110;

Qy    11 KAAVHTAAHNPVG---HG-VALQ--QGSSSSSPQN--------AAASLAAEGKNRGKMPR 56
            | :|  |  ||  :   :        ::|:  :         :::  |  | |  |
Db   428 KSTVLQQYNRVGKVEHGSVALPAIMRSGASGPENFQVGSMPPAQQQITSGQMRGHMPP 487

Qy    57 IHQPSTAADG------------ISAAHQQKKSFSLRGCLG---TKKFSRSAPQGQPGTTHSK- 103
            ||  ::         : :: | |       :: : :  : |  ||:
Db   488 LTSAQQALTGTINSSMQAVQAAQATLDDFETLPLGQDAASKAWRKNKMDESKHEIHSQV 547

Qy   104 ------GATLRDLLARDDGETOHE-----AAAPDAARLTR-SGGVK--RRNMDDMA--GRP 148
                  || ||| :   :       :      :   |::    :::::    ||:||    |
Db   548 DAITAGTASVVNLTAGDPAETDYTAVGCAVTTISSNLTEMSRGVVKLLAALLEDEGGNGRP 607

Qy   149 MVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQH---SPPHIPG 203
          :| :  || |  :  :::  :     :  :|      :|  |  :        |||| |
Db   608 LLQAAKGLAGAVSELLRSAQPASAEPRQNLLQAAGNVGQA-SGELLQQIGESDTDPHFQD 666

Qy   204 SHHEIKEPVGSTSKATTAHADRVEIAQEDDDSEFQQLHQQRLARERENPPQPPKLGVAT 263
            :   :|     || :  | :  |:    |:   :      :  :: ||  ||   | |
Db   667 VLMQL-ANAVASAAAALVLKAK--SVAQRTEDSGLQ---TQVIAAATQCALSTSQLVACT 720

Qy   264 PISARFQPKLTA-------VAESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAVT 310
          | :|  ||   |       | ::|  |  :::| |    ::   | |||  | | ||
Db   721 KVVA---PTISSPVCQEOLVEAGRLVAKAVEGCVSASQAATEDGQLLRGVGAAAT--AVT 775

Qy   311 LDKGKLQLAPDNPPALNTLLKQTLGKDTGHYLAHHASSDGSQHLLLDNKGHLFDIKSTAT 370
           | | ::: ||  | |  | |:| |               : |       :  ||  |: |
Db   776 ---QALNELL-----------QHVKAHATGAGPAGR------YD-QATDT 804

Qy   371 SYSVLHN-----SHPGEIKGK--LAQAGT--GSVSVDGKSGKISLGSGTQSHNKTMLS 419
             :|  |:      : |   |   ||| ::  ||  :   |  ::      |:|  :||:
Db   805 ILTVTENIFSSMGDAGEMVRQARILAQATSDLVNAIKADAE-----GESDLENSRKLLS 858

Qy   420 QP---GEAHRSLLTGIWOHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLS 476
          :      |||  |              || || :     |||       ::    :  |
Db   859 AAKILADATAKMV----EAAKGAA-------AHPD------SEEQQQRLR 891

Qy   477 RQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPS 536
           :| |   : : ::  ::|   :  ::|:  : :  :  : :  |  :  : |   |:  |
Db   892 EAAEG--LRMATNAAAQ----NAIKKKLVQRLE-HAAKQAAASATQTIAAAQHAAS-APK 943

Qy   537 LDASPESHISLSLH-FADAHQGLLHGKSELEAQSVAISHGRLVVADSE-----GKLFSA 589
            |: |   : :    || | | | : | |:|   : |  : :: :  :     | : :
Db   944 ASAGPQPLLVQSCKAVAEQIPLLVQGVRGSQAQPDSPSAQLAIIAASQSFLQPGGKMVAA 1003

Qy   590 A---IPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN------ 637
          |     | | || : || :    ||:    | ::     : ||   | :
Db  1004 AKASVPTIQD----QASAM-----------QLSQCAKNLGTALAELR 1035

Qy   638 -NFRQQHAC-PLGNDHQFHPGWNL--------TDALVIDNQLGLHHTNPEPHEIL------D 683
             :| :  |   |  |:|                |||: |:|  | |:|  | :       |
Db  1036 TAAQKAQEACGPLEMDSALSVVQNLEKDLQEIKAAARDGKL-----KPLPGETMEKCTQD 1090

Qy   684 MGH-------------LGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDG 731
           | |              |:||  | ::          | |  |  : | : | | | : |
Db  1091 LGNSTKAVSSAIAKLLGEIA--QGNENYAGIAARDVAGG---LRSLAQAARGVAALTSDP 1145

Qy   732 EVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYL-- 789
           | :  ::: :| : : :      | | ::  ||| | | |  || |||   :|  |
Db  1146 AVQAIVLDTASDVLDKASSLIEEAKKASGHPGDPESQQRLAQVAKAVTQA---LNRCVSC 1202

Qy   790 ---ALTEKGD------IRSFQIKPGTQQLERPAQTLSREGISG------ELKDIHV 830
             | ::| |       :|:  :  | :::   || :  :| |:       || : :
Db  1203 LPEGQRDVDNALRAVGDASKRLLSDLLPPSTGTFQE-AQSRLNEAAAGLNQAATELVQASR 1261

Qy   831 DHKQNL-YALTHEGEVFHQPREAWQNGAESSSWHKLALPQSE----SKLKSLDMSHEH 883
           :  ::  || :  |  :   | |  |   |   | |  | ::    ||||||| ||:|
Db  1262 GTPQDLARASGRFGQDFSTFLEA---GVEMAG-----QAPSQEDRAQVVSNLKGISMSSSK 1314

Qy   884 -----KPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTV-------------- 923
               | :|  |: : ||   :|   :|    : |  | |
Db  1315 LLLAAKALST--DPASPNLKS------QLAAAAR---AVTDSINQLITMCTQQAPGQKECDN 1365

Qy   924 -----------------------------FNRLMQGVKGKVIPGSGLT-VKLSAQTG----GMT 953
                                          :|:|:|| :| ::|  :| | :|: :| |
Db  1366 ALRQLETVRELLENPVQPINDMSYFGCLDSVMENSKVLGEAMTGISQNAKNGNLPEFGDA 1425

Qy   954 GAEGRKVSSKFSERIRAYAFNPTMSTPPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQ 1013
           ||   || |    |  ||    ::||  ||: | || | ||  |:  :  |  :|:|   :
Db  1426 IATASKALCGFTEAAAQAAYLVGVSDP-------NSQAGGQGLVEPTQFARA------ 1470

Qy  1014 LDAHNVRHNAPQPDLQSKLETLDGEHG-----AELLNDMKRFRDELEQSATRSVTVLGQHQ 1070
            : :|    | |   | :   |    :       : :  : :       |:  ::  |:| :
Db  1471 --NQAIQMACQS-----LGEPGCTQAQVLS-----AATIVAKHT 1502

Qy  1071 GVLKSNGEINS--EFKPSPGKALVQSF-NVNRSGQDLSKSLQ-------QAVHAT 1115
            :| |:   |:   | :   | :|||  :|:  |  |  | | :       ||  :|
Db  1503 SALCNSCRLASARTANPTAKRQFVQSAKEVANSTANLVKTIKALLDGFTENRAQCRAAT 1562

Qy  1116 PPSAES--KLQSMLGH--------FVSAGVDMSHQKGEIPLGR-- 1148
          | :|      :|::|         ||||| |::  :  | : |:
Db  1563 APLLEAVDNLSAFASNPEFSSVPAQISPEGRAAMEPIVISAKTMLESAGGLIQTARALAV 1622

Qy  1149 -QRDPNDKTALT-KSRLLDTV-------------- 1168
           :|||  :  |  ||  : |:
Db  1623 NPRDPPRWSVLAGHSRTVSDSIKKLLITSMRDKAPGQLECETAIAALNSCLRDLDQASLAA 1682

Qy  1169 -----------------TIGELHELADKAKLVSDHKPDADQ---IKQLRQFD- 1201
                              |  : | |:| | : :|  |: |    :|| ::|
Db  1683 VSQQLAPREGISQEALHTQMLTAVQEISHLIE-PLASAARAEASQLGHKVSQMAQYFEP 1740

Qy  1202 -TLREKRYESNPVKHYTDMGFTH--------NKALEANYDAVKAFINAFKKEHHGVNLTTRT 1254
           || :     :||  : :  | |         :|| ||   ||  :  :  : :  |  ::
Db  1741 LTLAAYGAASKTLSHPQQMALLDQTKTLAESALQLLYTAKEAGGNPKQAAH------TQE 1794

Qy  1255 VLES--QGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAG 1312
           : |   | |||   : :| | : ::|  | | :  :|| ::  :   :   | | ||: |
Db  1795 ALEEAVQMMTEAVEDLTTTL----NEAASAAGVVGGMVDSI-TQAINQLDEGPMGDPEG 1848

Qy  1313 ITLDRAYNLSFSRTSGGLNVSF----------GRDGGVSGNIMVATGHDVMPYMTGKKTS 1362
           : :|| |   |::     |: |           |:| | : :  : : : :    | :
Db  1849 SFVD--YQTTMVRTAKAIAVTVQEMVTKSNTSPEELGPLANQLTSDYGRLASQAKPAAVA 1906

Qy  1363 AGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPA 1422
          | |    : : :|  | | : :   |:|: ||| |:  :       || : | | ||| |
Db  1907 AENEEIGAHIKHRVQ-ELGHGCSALVTKAGALQCSPSDVYTKKEL-------IECA 1954

Qy  1423 ELLQKGIEH---QMKQGSKLTFSVDTSAN------LDLRAGI----NLNEDGSKP-- 1464
            : ::|     |::| :  |   : :: :      |||  |     :|| | | |
Db  1955 RRVSEKVSHVLAALQAGNRGTQACITAASAVSGIIADLDTTIMFATAGTLNREGAETFAD 2014

Qy  1465 -NGV--TARVSA-GLSASANLAAGSRERSTTSGQ-----------FGSTTSASNNR 1505
           :|   ||::   ||:|||: ||| ::|| ::| |            | ||:||: :
Db  2015 HREGILKTAKVLVEDTKVLVQNAAGSQEKLAQAQSSVATITRLADVKLGAASLGAEDP 2074

Qy  1506 PTFLNGVGAGANLTAALGVAHSTH--EGKPVGIFPAFTSTNVSAALALDNRTS------ 1557
           |  :|| ||| ||:| || |:||   |||| | |   || ||:: | | | ||
Db  2075 ETQVVLINAVKDVAKALGDLISATKAAAGK-VGDDPAVWQLKNSAKVMVTNVTSLLKTVK 2133

Qy  1558 ------------QSISLEL------KRAEPVTSNDISELTSTLGKHFKDSATTK 1593
                      |||:|:|      |:|| |:|: | | ||| | | |: |||:|
Db  2134 AVEDEATKGTRALEATTEHIRQELAVFCSEPPPAKTSTPEDFIRMTKGI----TMATAK 2188

Qy  1594 MLA-------ALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLV--- 1638
           ::        | ::  |:   :: :||: :  ||   :   : | ||: |  :
Db  2189 AVAAGNSCRQEDVIATANLSRRATADMLRACKEAAFHPEVAPDVRLRALHYGRECANGYL 2248

Qy  1639 -----------IRQQAADSHSMELGSASH---------------STTYNNLS 1664
                      ||:|| |||| || |:                :||  :
Db  2249 ELLDHVLTLQKPNPDLKQQLTGHSKRVAGSVTELIQAAEAMKGTEWVDPEDPTVIAENE 2308

Qy  1665 -----QAPSQEDRAQVVSNLKGISMSSSK 1724
Db  1665 RINNDGIVELLHKHFDAALPASSAKRLGEMNNDPALKDIIKQLQSTPFSSASVSMELKD 1724
```

```
Db   2309 LLGAAAAIEAAAKKLEQLKPRAKPKEADESLNFEEQILEAAKSI------AAATSALVKA 2362
          :|         ::|   |||:| ::                            :| :||:|
Qy   1725 GLREQTEKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQSVSKS---------EGFN 1772
          |:::|  |:||   ||||  ::::: :::                          :|
Db   2363 ASAAQRE-LVAQGKVGAIPANAL--DDGQWSQGLISAARMVAAATNNLCEAANAAVQGHA 2419
          ||  | |   ||| ||  :|:                ::|  ||| ::     | :|
Qy   1773 TPALLLGTSNSAAMSMERNIGTINFKYGQDQNTPRRFTLEG-GIAQAN------PQVASA 1825
          :|:||  :|::|   ||| : |||:  ||  |:::|    ::|           |:||:|
Db   2420 SQEKLISSAKQVAASTAQLLVACKVKADQDSEAMKRLQAAGNAVKVRASDNLVKAAQKAAA 2479

Qy   1826 LTDLKKEGLEMK 1837

Db   2480 FEDQENETVVVK 2491
```

```
RESULT
YD86_SCHPO   5
ID   YD86_SCHPO     STANDARD;     PRT;   1957 AA.
AC   Q10411;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   HYPOTHETICAL 222.8 KDA PROTEIN C1F3.06C IN CHROMOSOME I.
GN   SPAC1F3.06C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; Z70690; CAA94624.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE 1957 AA;  222785 MW;  3F480CA06171D9DA CRC64;
```

```
Query Match        2.3%;
Best Local Similarity  18.7%;       Score 217.5;  DB 1;  Length 1957;
Matches 386;    Conservative 298;   Mismatches 731;   Indels 645;   Gaps  93;
                Score 217.5;  Pred No. 0.012;

Qy     33 SSSSPQNAAASLAAEGKNRGKMPRIHQPST--AADGISAAHQKKSFSLRGCLGTKKFS 89
          ::||| |: || : ||||    |         |||||:|| ||:|::| || | |:
Db     18 SASSLVSSAASPFIDSLETPRPNISRASTGQLAEDGDTSSQHEDSSEELK-------- 68
```

```
Qy     90 RSAPGQGPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLT-RSGGVKRRNMDDMAGRP 148
          |  | | | | ||  :|  :|        || ::  | |     |  :||||:
Db     69 ---RQEVRGMRRHSDLSIDAKLGSSEGST-----ASSALPLTPRS--------PSNASWL 112

Qy    149 MVKGGSGEDKV----PTQQKRHOLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIPGS 204
          ::||          |     | | | || :: ::  |:              |
Db    113 LVRGGLLDSPILDINSVTQKSNLLNELKQVR-----SKLAALEHENGILSLIQ----LSSS 163

Qy    205 HHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ---LHQQRLARERENPPQPPKLG 260
          |   |: | |:|  :||:|         | : :         |: |::        ||
Db    164 NKKDK------NTSSVT+-----LTSEEDVSVFQKKLTMESNFSAKQSEAYDLSRQLL 211

Qy    261 VATPISARFQPKLTAVAESVLEGTDTTQSPLKFQSMLKKSGAGVTPLAVTLDKGKLQLAP 320
          :| :|: || |:||  :|| | | :|:| | ::|:  | ::|| :|||  || |:|:|
Db    212 TVTEKLDKKEKDYEKIKEDVSSIKASLAEEQASNKSLRGEQERLEKLLVSSNK------- 264
```

```
Db    321 DNPPALNTLLKQTLG----------KDTQHYLAHHASSDGSQOHLLLDNKGHLFDIKSTATSYS- 373
          ::|          | |          |::|:   :::|   ||::    :| | ::   :|
Qy    265 -----TVST-LRQTENSLRAECKTLQEKLEKCAINEEDSKL------LEELKHNVANYSD 312
```

```
Qy    374 -VLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGI 432
          ::|  |        | ||:| :  :  :::                           :|
Db    313 AIVH------KDKL---IEDLSTRISEFDNLKSERDT-LSIKNEKLEKLL-------- 352
```

```
Qy    433 WQHPAGAARPQGES-IRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYAL-KDNR 490
          :| :       ::|   |           | :|       |:|| : ::  | | |::
Db    353 -RNTIGSLKDSRTSNSQLEEMVEL-----------KESNRTIHSQLT-DAESKLSSFEQENK 402
```

```
Qy    491 TLQNLSDN---KSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHI 545
          :|       : :: :| ::|:  |: :|
Db    403 SLKGSIDEYQNNLSSKDKMVKQVSSQLEEARSSLAHAT--GK-------------- 442
```

```
Qy    546 SLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKA- 604
          :                    |       ::|   |: :::|||           |||::|
Db    443 ---------LAEINSERDFQNKKIKDFEKIEQDLRACLNSSS--------NELKEKSA 483
```

```
Qy    605 ------MPQHALDEHFGHDHQISGFFHDDHGQLNALVKN--NFRQQHACPLGNDHQFHPG 656
          |           |:   :   |:::     |:|        :: :|| ::| |
Db    484 LIDKKDQELNNLREQIKEQKKVS--ESTQSSLQSLQRDILNEKKKHE----------- 528
```

```
Qy    657 WNLTDALVIDNQL----GLHHTNPEPHEILDMGHLGS----LALQEGKLHYFDQLTKGW 707
          |::||           |  :|         || :|        :::   : ::||
Db    529 -------VYESQLNELKGELQT--EISNSEHLSSQLSTLAAEKEAAVATNNELSESK 576
```

```
Qy    708 TGAESDCKQLKKGLDGAAYLLKDGE------------VKRLNINQS---------TSSIKHGT 749
          :|   |                           ::||  :|           :|::|:|
Db    577 NSLQTLCNAFQEKLAKSVMQLKENEQNFSSLDTSFKLNESHQELENNHQITTKQLKDTS 636
```

```
Qy    750 ENVFSLPHVR-NKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQ--IKPGT 806
          ::|  |              | |:  |:|:: :|              :::    ||||
Db    637 SKLQQLQLERANFEQKESTLSDENNDLRTKLLKLEESNKSL-IKKQEDVDSLEKNIQTLK 695
```

```
Qy    807 QQLERPAQTL---------------------------SREGISGELKDIH----------- 829
          ::|    |:                          ||: |:|::|
Db    696 EDLRKSEEALRFSKLEAKNLREVIDNLKGKHETLEAQRNDLHSSLSDAKNTNAILSSELT 755
```

```
Qy    830 ---VDHKQ--NLYALTHEGEVFHQPREAWQNGAESSS--WHKL-----------ALPQ 869
          :|||      : ::| |:    ::: :: |  |::     :||:             :|
Db    756 KSSEDVKRLTANVETLTQDSKAMKQSFTSLVNSYQSISNLYHELRDHVNMQSQNNTLLE 815
```

```
Qy    870 SESKLKSLDMSHEHKPIATFEDGSOHQLKAGGWHAYAAPE--------------------- 909
          ||||| | :                                  ::| |:
Db    816 SESKLKT-DCENLTQQNMTLIDNVQKLM----HKHVNQESKVSELKEVNGKLSLDLKNL 869
```

```
Qy    910 RGPLAVGTSGSGQTVFNRLMQ----------------IPGSGLTVKL 945
          |||  |  |:  |:::                     :|
Db    870 RSSLNVAISDNDQIITQLAELSKNYDSLEQESAQLNSGLKSLEAEKQLLHTENEELHIRL 929
```

```
Qy    946 SAQTGGMTGAE------GRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREG 999
          :|  | |  |:              ||:: |:|  |  | :
Db    930 DKLTGKLKIEESKSSDLGKKLTARQEE---------------------------------- 956
```

```
Qy   1000 LKPLYEMQGALIKQLDAHNVRHNAPQPDLQSKL-ETLDLGEHGAELLNDMKRFRDELEQS 1058
          :        |:| :|   |::                      |  :|:|||: ::|
Db    957 ---------ISNLKEENMSQSQAITSVKSKLDETL--SKSSKLEADIEHLKNKVSEV 1002
```

```
Qy   1059 ATRSVTVLGQHQGV--LKSNGE--------------INSEFKPSPGKA 1090
          |:                ||::||               ::| | ||
Db   1003 EVERNALLASNERLMDDLKNNGENIASLQTEIEKKRAENDLQSKLSVVSSEYE----NL 1058
```

```
Qy   1091 LVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGV--DMSHQKGEIPLG 1147
          |:::|:    | ::|::    |     |      |:|:: |::||     |::  |  |
Db   1059 LLISSQTNKSLEDKTNQLKY-------IEKNVQKLLDEKQRNVELEELTSKYGKL--- 1107
```

```
Qy   1148 RQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQOFDTLREKR 1207
          :: :     | |   ||:| |::   |   | |::   ::|| :| :::     |:|:
Db   1108 -----GEENAQIKDELLALRKKSKKQHDLC--ANFVDDLKEKSDALEQLTNEKNELIVSL 1160
```

```
Qy   1208 YESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKK----EHHGVNLTTRTVLESQGSAE 1263
```

```
Db  1161  EQSN--------SNNEALVEERSDLANRLSDMKKSLSDSDNVISVIRSDLVRVNDELD 1210
              :||            :||||        ::       ::  ::
Qy  1264  LAKKLKNTLLS--------LDS-GESMSFSRSYGGVSTVFVPTLSK-KVPVPVI 1308
              || |::||       |||  |||||    |     ||| |||
Db  1211  TLKKDKDSLSTQYSEVCQDRDDLLDSLKGCEESFNKY---AVSLRELCTKSEIDVPVSEI 1267
Qy  1309  --------PGAGITLDRAYNLSFSRTSGGLN-VSFGR--DGGVSGNIMVATGHDVMPYM 1356
Db  1268  LDDNFVFNAGNFSELSRLTVLSLENYLDAFNQVNFKKMELD-----NRLTTTDAEFTKVV 1322
Qy  1357  TGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLK-FKLTEDELPGFIHGLT 1415
Db  1323  ADLEKLQHEHDDWLIQR-------GDLEKALKDSEKNFLRKEAEMTENIHSLE 1368
                                 ::
Qy  1416  HG--TLTPAELLQKGIEHQMKQGSKLTFSVDTSANLDLRAGINLNEDGSKPNGVTARVS 1472
              :|     :||   |   || ::|:  |        :|  :
Db  1369  EGKEETKKEIAELSSRLEDNQLATNKLNQLD-----HLNQEIRLKEDVLK----EKES 1418
Qy  1473  AGLSASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVG-AGANLTAALGVAHSSTHE 1531
Db  1419  LIISLEESL---SNQRQKES----SLLDAKNELEHMLDDTSRKNSSLMEKIESINSSLDD 1471
Qy  1532  GKPVGIFPAFTSTNVSAAL---ALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFK 1587
              |:|:|::|        ||        ::||| |  |:|||
Db  1472  --------KSFELASAVEKIGALQKLHSESLSL------MENIKSQLQEAKEKIQV 1513
Qy  1588  DSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSH 1647
Db  1514  DEST-----IQELDHEITASK------------NNYEGKLNDKDSIIRDLS----- 1547
Qy  1648  SMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQ 1707
Db  1548  ------------ENIEQLNN-----LL-----AEEKSAVKRLSTE----------KE 1572
Qy  1708  LQSTPFSSASVSMELKDGLREQTEKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQSVSK 1767
              ::       |::  ::|      |: ||:    ||:||:|:  ::
Db  1573  SEILQFNSRLADLEY-------HKSQVESELG-----RSKLKLASTTEELQLAE 1614
Qy  1768  SEGFNTPALLGTSNSAAMSMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQ------ 1821
Db  1615  NERLSLTTRMLDLQNQV----KDLSNIKDSLSEDLRTLR--SLEDSVASLQKECKIKSN 1667
Qy  1822  ---VASALTDLKKEGLEMK 1837
                   ||:|||||||:
Db  1668  TVESLQDVLTSVQARNAELE 1687
```

RESULT   6
AKAC_HUMAN
ID   AKAC_HUMAN   STANDARD;   PRT;   1781 AA.
AC   Q02952; Q99970; O00498; O00310;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   A-KINASE ANCHOR PROTEIN 12 (A-KINASE ANCHOR PROTEIN 250 KDA) (AKAP
DE   250) (MYASTHENIA GRAVIS AUTOANTIGEN GRAVIN).
GN   AKAP12 OR AKAP250.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Heart;
RX   MEDLINE=97153077; PubMed=9000000;
RA   Nauert J.B., Klauck T.M., Langeberg L.K., Scott J.D.;
RT   "Gravin, an autoantigen recognized by serum from myasthenia gravis
RT   patients, is a kinase scaffold protein.";
RL   Curr. Biol. 7:52-62(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Umbilical vein endothelial cells;

RX   MEDLINE=98269042; PubMed=9604001;
RA   Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;
RT   "Changes of gene expression by lysophosphatidylcholine in vascular
RT   endothelial cells: 12 up-regulated distinct genes including 5 cell
RT   growth-related, 3 thrombosis-related, and 4 others.";
RL   J. Biochem. 123:1119-1126(1998).
RN   [3]
RP   SEQUENCE OF 43-1781 FROM N.A.
RC   TISSUE=Umbilical vein endothelial cells;
RA   Bowditch R.D., Ginsberg M.H.;
RT   "Sequence of gravin cDNA isolated from a human umbilical vein
RT   endothelial cell library.";
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1477-1781 FROM N.A.
RC   TISSUE=Umbilical vein endothelial cells;
RX   MEDLINE=92395179; PubMed=1522245;
RA   Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.,
RA   Lindstrom J., Ginsberg M.H.;
RT   "Molecular cloning and preliminary characterization of a novel
RT   cytoplasmic antigen recognized by myasthenia gravis sera.";
RL   J. Clin. Invest. 90:992-999(1992)
CC   -!- FUNCTION: ANCHORING PROTEIN THAT MEDIATES THE SUBCELLULAR
CC   COMPARTMENTATION OF PROTEIN KINASE (PKA) AND PROTEIN KINASE C
CC   (PKC).
CC   -!- SUBUNIT: BINDS TO DIMERIC RII-ALPHA REGULATORY SUBUNIT OF PKA.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY BE PART OF THE CORTICAL
CC   CYTOSKELETON.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; MAY BE
CC   PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN ENDOTHELIAL CELLS, CULTURED
CC   FIBROBLASTS AND OSTEOSARCOMA, BUT NOT IN PLATELETS, LEUKOCYTES,
CC   MONOCYTIC CELL LINES OR PERIPHERICAL BLOOD CELLS.
CC   -!- INDUCTION: ACTIVATED BY LYSOPHOSPHATIDYLCHOLINE (LYSOPC).
CC   -!- DOMAIN: POLYBASIC REGIONS LOCATED BETWEEN RESIDUES 265 AND 556 ARE
CC   INVOLVED IN BINDING PKC.
CC   -!- DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY
CC   PATIENTS WITH MYASTHENIA GRAVIS (MG).
CC   --
DR   EMBL; U81607; AAC51366.1; -.
DR   EMBL; AF001504; AAB58938.1; -.
DR   EMBL; AB003476; BAA19927.1; -.
DR   EMBL; M96322; AAA35931.1; -.
DR   MIM; 604698; -.
DR   PIR; A43922; A43922.
DR   InterPro; IPR001573; -.
DR   InterPro; IPR001899; -.
KW   Antigen; Alternative splicing.
FT   DOMAIN   603   633   AKAP 1.
FT   DOMAIN   752   782   AKAP 2.
FT   DOMAIN   797   827   AKAP 3.
FT   DOMAIN   101   101   POLY-GLU.
FT   DOMAIN   265   556   INVOLVED IN PKC-BINDING (PROBABLE).
FT   DOMAIN   1540   1553   RII-BINDING (PROBABLE).
FT   VARSPLIC   1   98   MISSING (IN ISOFORM 2).
FT   VARSPLIC   99   106   EEEVVTE -> MLGTITIT (IN ISOFORM 2).
FT   CONFLICT   117   117   E -> K (IN REF. 2).
FT   CONFLICT   142   144   NRN -> TPEI (IN REF. 2 AND 3).
FT   CONFLICT   215   215   Q -> K (IN REF. 2).
FT   CONFLICT   448   448   G -> E (IN REF. 2 AND 3).
FT   CONFLICT   694   694   R -> G (IN REF. 2 AND 3).
FT   CONFLICT   867   867   G -> S (IN REF. 2 AND 3).
FT   CONFLICT   986   986   S -> A (IN REF. 2 AND 3).
FT   CONFLICT   1530   1530   E -> EE (IN REF. 3 AND 4).
FT   CONFLICT   1581   1581   V -> M (IN REF. 4).

```
FT   CONFLICT    1601   1601   Q -> L (IN REF. 2).
SQ   SEQUENCE    1781 AA;  191439 MW;  BA81393737 9FAC0F CRC64;

Query Match        2.38;   Score 214;   DB 1;   Length 1781;
Best Local Similarity   17.9%;   Pred. No. 0.015;
Matches  339;  Conservative 278;  Mismatches 718;  Indels  564;  Gaps   80;

Qy     29  QQGSSSSSPQNAAASLAAEGKNRGKMPRIHQPSTAADG------ISAAHQQKKSFSLRGCL   83
            : :::::::::       :  :: :: :: :: ::   ::    ::    : ::: ::
Db    245  EQSHAEISPPAESGQAVECKEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQGWA  304

Qy     84  GTKK---FSRSAPQGQPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLTRSGGVKRRN  140
            : ::    :   ::  :  :  ::  : ::  :: :  :::  : :  ::::::
Db    305  GWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGK-----AEVASEKLTASEQAHPQE  359

Qy    141  MDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPH  200
            ::  ::::         :: ::::
Db    360  PAESAHEPRLSAEYEKVELPSEEQ-----------------------------------  383

Qy    201  IPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDSEF-------QQLHQQRLARERE--N  252
            ::      :::                 :::        : :::  :: :::::
Db    384  VSGSQGPSEEKP---APLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQKTEVEETAGS  440

Qy    253  PPQPPKLGV-ATPISARFQPKLTAVAESVLEGTDTTQ-SPLKP-QSMLKGSGAGV---TP  306
            :   :  :       :  : ::  ::    :: ::     ::   :     ::     ::
Db    441  VPAEELVGMDAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVE  500

Qy    307  LAVTLDKGKLQLAPDNPPALNTLLKQTIGK-----------DTQHYLAHHASSDGSQHLL  355
            :: : :::    ::::  : :  :  ::              ::   : :::  :: :
Db    501  MLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKRGGDEESGEHTQVPADSPDSQE--  558

Qy    356  LDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQEAGTGSVSVDGKSGKISLGSGTQSHNK  415
            :: ::: : :  : ::: :::          : ::::    :: ::      : :  :
Db    559  EQKG-------ESSASS------PEEPEEI--TCLEKGLAEVQQDGEAEE---GATSDGEKK  602

Qy    416  TMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQL  475
            :  :::::  : : :: :    ::::: ::::     : : :   : :  :: :: :::
Db    603  REGVTPWASFKKMVT----PKKRVRRPSES------DKEDELDRVKSAATLSSTESTASEM  652

Qy    476  SRQADGKLYALKDNRTLQNLSDNKSSEKLV------DKIKSYSVDQRGQVAI-----L  522
            :  :: : : :    :: ::  :: :: :        :  : :: ::   :      :
Db    653  QEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRRSSSDEEGGPKAMGGDHQK  712

Qy    523  TDTPGRHKMSIMPSLDASPESHISLSLHFADAHQG----LLHGKSELEAQSVAISHGRL  577
            :  ::     ::::  :: : :  :: :           :::    ::  :: :: :
Db    713  ADEAGKDKETGTDGILAGSQEH------DPGQGSSSPEQAGSPTEGEGVSTWESFKRL  764

Qy    578  VVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN  637
            :  :: ::: :     ::      : :   :: ::    ::::   :: :   ::::
Db    765  VT--------PRKKSKSLEEKS-----------EDSIAG------------------785

Qy    638  NFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKL  697
            :: :     :::  :  :: :  ::  :  :: :    : :  ::   ::: :   ::
Db    786  ---------------------------------SGVEHSTPDTE--------796

Qy    698  HYFDQLTKGWTGAESDCKQLKKGLDGAYLLKDGEVKRLNIQSTSSIKHGTENVFSLPH  757
            :  :: ::     :: : :::   : :  ::  :: ::   ::: :  :::  :: :
Db    797  -------PGKEESWVSIKKFIPGRRKKRPDGKQEQAPVEDAGPT----------833

Qy    758  VRNKPEPGDALQLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLS  817
            : ::     :  ::::  :::  :    :  ::   :: : :  :: ::: ::
Db    834  ------------GANEDD-SDVPAVVPLSEYDAV-EREKMEAQQAQKGAEQPEQKAAT--877

Qy    818  REGISGELKDIHVDHKQNLYA-LTHEGEVFHQPREAWQNGAESSSWHKL--ALPQSESK  873
            :::      ::  :: :::       :::  :  :   :::   :        :: :::
Db    878  --EVSKELSESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEV  935

Qy    874  LKSLDMSHEHKPIATFEDGSQHLKACGWHAYAAPERGPLAVCTSGSQTVFNRLMQGVKG  933
            :: : :: :  ::  :: :: :::   :: :  ::   :  ::: : :::::: :: :
Db    936  LEREVIAEEEPPTVT----------EPLPE---------NREARG--D 962

Qy    934  KVIPGSGLTVK--LSAQTGGMTGA-EGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYAT 990
```

```
Db    963  TVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETEMVSAVSQLITDSPDTEEAT---  1019
            ::  :: :                     : :::::    :::
Qy    991  QHGWQGREGLKPLYEMQGAL-----------IKQLDAHNVRHNAPQPDLQSKLETLDLG  1038

Db   1020  --------------PVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLP--------G 1057
                          :::::  ::  ::: :: ::: :::      ::
Qy   1039  EHGAE-LLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSP---GKALVQS 1094

Db   1058  TGGPEDVLQPVQRAEAERPEQAEASGLKKETDVVLKVDAQ--EAKTEPFTQGKVV----1111
            :::  : :: :::  ::  ::  :  :::    ::  ::    : : :: :
Qy   1095  FNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPND 1154

Db   1112  -----GQTTPESFEKAPQVTESIESSELVTTCQAETLAGV----KSQEMVMEQAIPPD 1160
                 :  : :  ::  : :: ::  :: ::  ::      :  :  ::  :
Qy   1155  KTALTKSRLLLDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVK 1214

Db   1161  S---------VETPTDSETDGSTPVADFDAPGTTQKDEIVEI-----------HEENEVA 1200
                     ::   : ::::  :        :: :::
Qy   1215  HYTDMGFTHNKALEANYD---AVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNT 1271

Db   1201  SGTQSGGTEAEAVPAQKERPPAPSSFV-FQEE----------TKEQSKMEDT 1241
            :: :::::: :::      :: :::  :             :: :: :
Qy   1272  LLSLD--SGESMS-FSRSYGGVGVSTVFVPTLSKKVPVPVIPGAGITLDRAVNLSFSRTS 1327

Db   1242  LEHTDKEVSVETVSILSKTEGTQEADQYADEKTKDVPF---------------FE----1281
            :: ::  : : ::    :::   : ::  ::  :: ::               ::
Qy   1328  GGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNA--SDWLSAK--HKISPDLRIGA 1384

Db   1282  --------GLEGSIDTGITVSREKVTEVALKGEGTEEAECKKDDALELQSHAKSP-----1328
                    : :  ::: :::   ::  :   : :::  :::  :: :: :
Qy   1385  AVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVD 1444

Db   1329  --PSPVEREMVVQVEREKTEAE----------------PTHVNEEKLEHE---TAVTVSEE 1368
              : : ::  : :::::  ::                : :::: ::: :       :: :
Qy   1445  TSANLDLRAGINLNE--------------DGSKP---------NGVTARVSAGLS-ASANLAAGSRERS 1489

Db   1369  VSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEK 1428
            :: : :  : :   :   :  :    ::    ::::  : ::  :::  : : :::::::
Qy   1490  TTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHS----STHEGK--PVG-------1536

Db   1429  V----LGETANILETGETL-----EPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAK 1481
                 : :: :: :::        :: :::: :::  :::::  ::      ::  ::::
Qy   1537  ----IFPAFTSTNVSAALALDNRTS------------QSISLELKRAEPVTSNDIS 1576

Db   1482  STPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEVKVSVAIEDLEP--ENGIL 1539
            :: : :::   :::: :: : :  ::::  ::: : :: ::: :: : : : :   : ::
Qy   1577  ELTSTLGKHFK----------DSATTKMLAALKELDDAKPAEQLHIL---QQHFSA 1619

Db   1540  ELETKSSKLVQNIIQTAVDQFVRTEETATEMLTS--ELQ----TQAHVIKADSQDAGQE 1592
            :: ::: ::: :: ::: ::: : ::: :::     :::      ::   ::    ::
Qy   1620  KDVVGDERYEAVRNLKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHF 1679

Db   1593  TEKEGEEPQASAQD--ETPITSAKEESESTAVGQA-HSDISKDMS-------EASEKTM 1641
            :: :::::  ::     :: ::::: :::: :::  ::  ::         :: :::
Qy   1680  DAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDG---LREQTEKAILD 1736

Db   1642  TVEVEGSTV-------NDQQLEEVVLPSEEGGGAGTKSVPEDDGHALLAERIEKSLVE 1693
            :: : ::         :  :: :: :: ::    ::::  :: :: :: ::  :::
Qy   1737  GKVGREEVGVLFQDRNNLRVKSVSVSQSVSK-SEGFNTP 1774

Db   1694  PKEDEKGDDVDDPENQNSALADTDASGGLTKESPDTNGP 1732
```

```
RESULT   7
FHAB_BORPE
ID   FHAB_BORPE     STANDARD;     PRT;   3591 AA.
AC   P12255;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   FILAMENTOUS HEMAGGLUTININ.
GN   FHAB.
```

```
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=90355839; PubMed=2388559;
RA   Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT   "Genetic characterization of Bordetella pertussis filamentous
RT   haemagglutinin: a protein processed from an unusually large
RT   precursor.";
RL   Mol. Microbiol. 4:787-800(1990).
RN   [2]
RP   SEQUENCE OF 1-3261 FROM N.A.
RX   MEDLINE=89202384; PubMed=2539596;
RA   Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;
RT   "Filamentous hemagglutinin of Bordetella pertussis: nucleotide
RT   sequence and crucial role in adherence.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641(1989).
-!- FUNCTION: EVIDENCE FOR A ROLE IN HOST-CELL BINDING AND
    INFECTION.
-!- SUBCELLULAR LOCATION: SURFACE.
```

```
DR   EMBL; M60351; AAA22974.1; -.
DR   EMBL; M60351; AAA22975.1; ALT_INIT.
DR   EMBL; M60351; AAA22976.1; ALT_INIT.
KW   Antigen; Hemagglutinin.
SQ   SEQUENCE   3591 AA;   367420 MW;   EF7418B30D6E5138 CRC64;
```

```
Query Match     2.2%;     Score 207.5; DB 1;  Length 3591;
Best Local Similarity   19.3%;     Pred: No. 0.09;
Matches  427;    Conservative 268;    Mismatches 792;    Indels 725;   Gaps 114;
```

```
Qy     5  SLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQPSTAA   64
              : : ||  | |   |  ||  | :::|  |  ||| |||||     |
Db  1379  SLDIKKGGAQVTVAGRYAEHGEVSIQGDYTVSAD--AIALAAQVTQRG----------GA 1426

Qy    65  DGISAAHQQKKSFSLRGCLGTKKFSRSRSAPQGGQPGTTH-SKGATLRDLLARDDGETQHEAA  123
              : | |            |||  | ::| |  ::
Db  1427  ANLTSRHDTRFSNKIR-LMGPLQVNAGGPVSNTGNLKVREGVTV--TAASFDNETGAEVM 1483

Qy   124  APDAARLTRSGGVKRRNMDDMAGRPMVK-------------GGSG  155
            : : || |   ||||  | :
Db  1484  A--KSATLTTSGAARN-----AGKMQVKEAATIVAASVSNPGTFTAGKDITVTSRGGFDN 1536

Qy   156  EDKVPTQQ----KRHQLNN-------FGQMRQTMLSKMAHPASANAGDRLQHS  197
           : :| | ||    | :|  | :: ::  |
Db  1537  EGKMESNKDIVIKTEQFSNGRVLDAKHDLTVTASQQADNRGSLKAGHDFTVQA-QRIDNS 1595

Qy   198  PPHIPGSHHEIKEPVGSTSKATTAHADRVEIAQEDDDSEFQQLHQQRLARERENPPOPP  257
               :         :|   |  ||   :|
Db  1596  GTMAAGHDATLKAPHLRNTGQVVAGH----------DIHIINSAKLENTGRVDA 1639

Qy   258  KLGVATPISARFQPKLTAVAE-----SVLEGTD--TTQSPLKP--QSMLKGSGAGVTP  306
           : :|| | :   |  :| |  ||    :| |
Db  1640  RNDIALDV-ADFNTGSLYAEHDATLTLAQGTQRDLVVDQDHILFVAEGTLRVKAKSLTT 1698

Qy   307  LAVTLDRKGKL----QLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHL  362
            : |||  |   | |   | :|   ::   :
Db  1699  EIETGNPGSLIAEVQENIDNKQAI------VVGKDLTLSSA-HGNVANEANALLWAAGEL 1751

Qy   363  FDIKSTATSYSVLHNSHPGEIKGKLAQAG-----TGSVSVDGKSGKISLG-----SGTQS  412
            : :| |   | :    | : | :: ||     |  |  :| | :|
Db  1752  -----TVKAQNITNK------RAALIEAGGNARLTAAVALLNKLGRIRAGEDMHLDAPRI 1800

Qy   413  HNKTMLSQPGEAHRS------------------------------LLTGIWOHPAGAARP-----  442
            :    | |  :
Db  1801  ENTAKLS--GEVQRKGVQDVGGGEHGRWSGIGYVNYWLRAGNGKKAGTIAAPWTGGDLTA 1858

Qy   443  ---------QGESIRLH---DDKIHILHPELGVWQSADKDTHSQLSRQADGK---------  482
                 ||:|  |   |::  || |    :|  |
Db  1859  EQSLIEVGKDLYLNAGARKDEHRHLLNE--GVIQAGG--HGHIGGDVDNRSVVRTVSAM 1913

Qy   483  ----------LYALKDNRTLQNLSDN-KSSEKLVDKIKSYSVDQRGQVAILTDPGRHKM  531
                 | ||||| ||   :||  |  :    : ||
Db  1914  EYFKTPLPVSLTALDNRAGLSPATWNFQSTYELLD----YLLDQNRYEYIWGLYPYTEW 1969

Qy   532  SI------------------MPSLDASPESHISLSLHFADAHQG------LLHG---KS  563
            :                  : :| : : : : ||   |        ||| :
Db  1970  SVNTLKNLDLGYQAKPAPTAPPMPKAPE--LDLRGHTLESAEGRKIFGEYKKLQGEYEKA 2027

Qy   564  ELEAQSV-AISHGRLVVADSEGKLFSAAI------PKQGDGNELKMKAMPQHALDEHFGH  616
            : | |  | || | | :  |   |    |
Db  2028  KMAVQAVEAYGEATRRVHDQLGQRYGKALGGMDAETKEVDG------------IIQEFAA 2075

Qy   617  DHQISGFFHDDHGQLNA---LVKNNFRQQ-HACPLGNDHQFHPGWNLTDALVIDNQLG--  670
            :   :         :|     : :     |  |  :    | |   :
Db  2076  DLRTVYAKQADQATIDAETDKVAQRYKSOIDAVRL--QAIQPG-RVTLAKALSAALGAD 2131

Qy   671  --LHHTN----------------PEPHEILDMGHLGSLALQEGKLHYFDQLTKGW  707
            : :|||                  : : :| | | |  | |  | | | ||
Db  2132  WRALGHSQLMQRWKDFKAGKRGAEIAFYPKEQTVLAAG--AGLTLSNGAIH------NGE 2183

Qy   708  TGAESDCKQ-LKKG------LDGAAYLLKD-GEVKRLNINQSTSSIKHGTENVFSLPHV  758
             : | :  :  |:|       || |  | :  | |||  :   |:   :   : | | |
Db  2184  NAAQNRGRPEGLKIGAHSATSVSGSFDALRDVGLEKRLDIDDALAA------VLVNPHI 2236

Qy   759  ----------------RNKPE--------------------PGDALQGLNKDDKAQAMAVI  783
                            | |                     | :|  | | : | : | |
Db  2237  FTRIGAAQTSLADGAAGPALARQARQAPETDGMVDARGLGSADALASLDAAQGLEVS 2296

Qy   784  GVNKYLALTEKGDIRSFQIKPGTQQLERP--AQTLSREGISGELKD--IHVDHKQNLY  837
            | :: |  | :    :   | |: :|      | :| : | : | |   : |:|  :
Db  2297  G-RRNAQVADAG------LAGPSAVAAPAVGAADVGVEPVTGDQVDQPVVAVGLEQPVA 2348

Qy   838  ALTHEGEVFHQPREAWQNGAESSSWHKLALPQOSESKLKSLDMSHEHKPIATFEDGSOHQL  897
            : |  : :  |  : :|  ::    |  | |       :   :
Db  2349  TVRVAPPAVALPR-----------PLFETRIKFIDQSKFY------GSRYFF 2383

Qy   898  KAGGWHAYAAPERGPLAVGTSGSQTVF----NRLMQGVKGKV-IPGSGLTVKLSAQTGG  951
             : |:  | |  || |  |  :  : ||     | ||| : ||  :| | | || | ||
Db  2384  EQIGY-----KPDRAARVAGDNYFDTTLVREQVRRALGGYESRLPVRGVALVAKLMDSAGT 2439

Qy   952  MTGAEGRKV-------SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLY 1004
            : | | |:        |::     |  | : |      |:  |  | |||  | :| |
Db  2440  VGKALGLKVGVAPTAQQLKQADRDFVWYVDTVIDGQKVLAPRLYLTE--ATRQGITDQY 2496

Qy  1005  EMQGALIK-------QLDAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQ 1057
            : || |::        | | |   | |   |::  ::  | | : | :| | : : ::|
Db  2497  AGGGALIASGGDVTVNTDGHDV--SSVNGLIQGRSVKVDAGK-GKVVVADSKGAGGGIE- 2552

Qy  1058  SATRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPP 1117
            : :   :: |:  |:        :|      |:| |  : : :  |
Db  2553  -ADDEVDVSGRDIGI--EGGKLR------GKDVRLKADTVKVATSM 2589

Qy  1118  SAESKLQSMLGHFVSAGVD-MSHQKGEIPLGRQRDPNDKTALT--KSRLILDTVTIGELH 1174
             : :| :| :| :|||| | :|  :|   | :| :  |  | |   |: | | ||| |::
Db  2590  RYDDK-----GRLAARGDGALDAQGGGLHIEAKRLETAGATLKGGKVKLDVDVKLGGVY 2644

Qy  1175  ELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTH--NKALEANY 1231
            | |   ::    ||       : :|  |   | |:||   :    |     :|:
Db  2645  EAGSS-------YBN----KSSTPLGSLFAILSSTTETNQ 2673

Qy  1232  DAVKAFINAFKKEHHGVNLTTRTV--LESQQSAELAKKLKNTLLSLDSGESMSFSR 1285
            ::    :| | :|  ||     |   || |:|  | |:| |   |  :|   :|
Db  2674  SA------HANHYGTRIEAGTLEGKMQNLEIEGGSVDA--AHTDLSVARDARFKAAA 2722
```

```
Qy  1286  SYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTS-GLNVSFGRDGGVSGNI  1344
                |  |  ||   |||         |||        |   |:
Db  2723  DFAHAEHEKDVRQLSLGAKV----GAG---GYEAGFSLGSESGLEAHAGR--GMTAGA   2771

Qy  1345  MVATGHDVMPYMT-----------------GKKTSAGNASDWLSAKHKISPDL--RIGAAV  1386
            |  :               |||   |       ||  |
Db  2772  EVKVGYRASHEQSETEKSYRNANLNFGGGSVEAGNVLDIGGA------DINRNRYGGAA   2825

Qy  1387  SGTLQGTIQNSLKFKLTEDELPGFIHGLT----------HGTLTPAELLQKGI         1429
                ||: |    |||    |||
Db  2826  KGN-AGT-EEALRMRAKKVESTKYVSEQTSQSSGWSVEVASTASARSSLLTAATRLGDSV   2883

Qy  1430  EHQMKQGSKL------------TFSVDTSANLDLRAGINLNEDGSKPNGVTA-----     1469
            :  | :        ||: | ||: |
Db  2884  AQNVEDGREIRGELMAAQVAAEATQLVTADTAA-VALSAGISADFDSSHSRSTSQNTQYL   2942

Qy  1470  ------------------RVSAGLSASANLAAGSR-ERSTTSGQFGSTTSASN         1503
                              :| ||  ::|| |       :|  ||
Db  2943  GGNLSIEATEGDATLVGAKFGGGDQVSLKAAKSVNLMAAESTFESYSESHNFHASADANL   3002

Qy  1504  NRPTFLNGVGAGANLTAALGVAHSSTHE-GKPVGIFPAFTSTNVSAA-LALDNRTSQSIS  1561
            |||: |||  |:|    |::|: |  ::| :: ||
Db  3003  GANAVQGAVGLG--LTAGMGTSHQITNETGK-------TYAGTSVDAANVSID--AGKDLN  3052

Qy  1562  LELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAKD  1621
            |:        :::         ||:
Db  3053  LSGSRVR------------GKHVVLDVEGDINATSKQ------------            3077

Qy  1622  VVGDER-YE----------AVRNLKKLVIRQQAADSHSMELGSA--SHSTTYNNLSRI    1666
              ||| :           ||||          |:  | || :: :||:
Db  3078  ---DERNYNSSGGGWDASAGVAIQN-RTIV-------APVGSAGFNFNTEHDN-SRL     3122

Qy  1667  NNDGIVELL------HKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVS   1719
             |||  :::        | ::           ||          |:       ||:
Db  3123  TNDGAAGVVASDGLTGHVKGDANLTGATIADLSGKGN---LK------YDGAVNA       3168

Qy  1720  MELKDGLREQTEKAILDGKVGREEVGLVFQDRNNLRVKSVSVSQVSKSEGFNTPALLLG   1779
               | :         |||: |||          | |          || |  ||:   |||
Db  3169  QNLKD-YRDK------DGGSGGLNVG-------ISSTTLAPTVGVAFG-----RVAG     3206

Qy  1780  TSNSAAMSMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASALTDLKK          1831
              ::       |  |::          ||  ||| | | |  :|  | ::
Db  3207  EDYQAEQRATIDVG-------QTKDPARLQVGGGVKGTLNQDAAQAITVQR           3250
```

```
RESULT 8
APC_MOUSE
ID   APC_MOUSE        STANDARD;      PRT;    2845 AA.
AC   Q61315; Q62044;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN) (MAPC).
GN   APC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1 AND 2), AND VARIANTS.
RC   STRAIN=C57BL/6J, AND CAST/EI; TISSUE=Brain;
RX   MEDLINE=92263101; PubMed=1350108;
RA   Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA   Luongo C., Gould K.A., Dove W.F.;
RT   "Multiple intestinal neoplasia caused by a mutation in the murine
RT   homolog of the APC gene.";
RL   Science 256:668-670(1992).
RN   [2]
RP   ERRATUM.
RA   Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA   Luongo C., Gould K.A., Dove W.F.;
RL   Science 256:1114-1114(1992).
RN   [3]
RP   SEQUENCE OF 1-45 FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Liver;
RA   Dicker F., Lambertz S., Reitmair A., Ballhausen W.G.;
RT   "The murine APC gene: alternative splicing of 5' untranslated
RT   region segments.";
RL   Submitted (OCT-1993) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=94061824; PubMed=8242607;
RA   Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
RT   "APC gene messenger RNA: novel isoforms that lack exon 7.";
RL   Cancer Res. 53:5589-5591(1993).
CC   -!- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
CC       CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
CC       ALLOWS THE DOWNREGULATION OF CYTOPLASMIC BETA-CATENIN (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: FORMS HOMOOLIGOMERS AND ASSOCIATES WITH CATENINS (BY
CC       SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, SPLEEN, KIDNEY, HEART,
CC       LUNG, BRAIN, STOMACH, INTESTINE, TESTIS AND OVARY.
CC   -!- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M88127; AAB59632.1; -.
DR   EMBL; U02937; AAA03443.1; -.
DR   HSSP; Q02248; 2BCT.
DR   MGD; MGI:88039; Apc.
DR   InterPro; IPR000225; -.
DR   Pfam; PF00514; Armadillo_seg; 4.
DR   PROSITE; PS50176; ARM_REPEAT; 1.
KW   Anti-oncogene; Phosphorylation; Alternative splicing; Repeat;
KW   Coiled coil.
FT   DOMAIN         1     61        COILED COIL (POTENTIAL).
FT   DOMAIN       125    245        COILED COIL (POTENTIAL).
FT   DOMAIN         1    728        LEU-RICH.
FT   REPEAT       451    493        ARM 1.
FT   REPEAT       503    545        ARM 2.
FT   REPEAT       546    589        ARM 3.
FT   REPEAT       590    636        ARM 4.
FT   REPEAT       637    681        ARM 5.
FT   REPEAT       682    723        ARM 6.
FT   REPEAT       724    765        ARM 7.
FT   DOMAIN       739   2834        SER-RICH.
FT   DOMAIN      1130   1156        ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      1556   1575        ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      1864   1891        HIGHLY CHARGED.
FT   VARSPLIC     243    276        MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT   VARSPLIC     310    410        MISSING (IN ISOFORM 3 AND ISOFORM 4).
FT   VARIANT      120    120        T -> A (IN STRAIN CAST/EI).
FT   VARIANT      493    493        V -> I (IN STRAIN CAST/EI).
FT   VARIANT      797    797        Y -> F (IN STRAIN CAST/EI).
FT   VARIANT     1330   1330        A -> T (IN STRAIN CAST/EI).
FT   VARIANT     1618   1618        A -> T (IN STRAIN CAST/EI).
FT   VARIANT     2294   2294        G -> A (IN STRAIN CAST/EI).
FT   VARIANT     2496   2496        H -> Q (IN STRAIN CAST/EI).
FT   VARIANT     2523   2523        T -> A (IN STRAIN CAST/EI).
FT   VARIANT     2813   2813        T -> S (IN STRAIN CAST/EI).
SQ   SEQUENCE   2845 AA;  311086 MW;  145CA73CF570A499 CRC64.

Query Match             2.2%;  Score 204;  DB 1;  Length 2845;
Best Local Similarity  19.9%;  Pred. No. 0.093;
Matches  412;  Conservative  272;  Mismatches  808;  Indels  580;  Gaps  97;
```

```
Qy 1004 YEMQGALIKQLD----AHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSA 1059
Db 1100 YRSRGTSGSETNRMGSSHAINQNVNQSLCQEDDYEDDKPTNYSERYSEEEQHEEEEERPT 1159

Qy 1060 TRSVTVLGQHQGV-------LKSNGEINSEFKPSPGKALVQSFNVNRSGQD-----LSKS 1107
Db 1160 NYSIKYNEEKHHVDQPIDYSLKYATDISSSQKPS------FSFSKNSSAQSTKPEHLSPS 1213

Qy 1108 LQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKG--EIPLGRQRDPNDKTALTKSRLI 1164
Db 1214 SENT--AVPPS-NAKRQNQL-RPSSAQRNGQTQKGTTCKVP--------SINQETIQT-YCV 1262

Qy 1165 LDT-VTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTH 1223
Db 1263 EDTPICFSRCSSLSLSS--ADDEIGCDQTTQEADSANTLQTAEVKENDVTRSAEDPATE 1320

Qy 1224 NKALEANYDAVKAFINA-------FKKEHHGVNLTTRTVLESQGSAELAKK-----LKNTL 1272
Db 1321 VPAVSQNARAKPSRLQASGLSSESTRHNKAVEFSSGAKSPKSGAQTPKSPPEHYVQETP 1380

Qy 1273 L-------SLDSGESMSFSRSY----GGGVSTVFVPT-------- 1298
Db 1381 LVFSRCTSVSSLDSFESRSIASSVQSEPCSGMVSGIISPSDLPDSPGQTMPPSRSKTPPP 1440

Qy 1299 ------LSKKVPVPVIPGAGITLDRAYNLSFSRTSG----GLNVSFGRDGGVSGNIMVAT 1348
Db 1441 PPQTVQAKREVPKSKVPAAE-------KRESGPKQTAVNAAVQR------VQVLP 1482

Qy 1349 GHDVMPYMTGKKTSAG-NASDWLSAKHKISP-------DLRIGAAVSGTLQGTLQNSLKFK 1401
Db 1483 DVDTLLHFATESTPDGFSCSSSLSALSLDEPFIQKDVELRIMPPVQ------- 1528

Qy 1402 LTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFS---VDTSANLDL------ 1451
Db 1529 ---END--------NGNETESEQPEESNENQDKEVKPDSEKDLLDDSDDDDIEILEECI 1577

Qy 1452 ------RAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTSAS 1502
Db 1578 ISAMPTKSSRKAKKLAQTASKLPPVARKPSQLPVYKLLPAQNRLQAQKHVSF------- 1630

Qy 1503 NNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSISL 1562
Db 1631 ---------TPGDDVPRVYCVEGTPI--------NFSTATSLSDLTIESPPN 1665

Qy 1563 ELKRAEPVTSN-------DISELTSTLGKHFKDSATTKMLAAL-KELDDAKPAEQLHILQQ 1615
Db 1666 ELATGDGVRAGIQSGEFEKRDTIPTEGRSTDDAQRGKISSIVTPDLLDDNK-AEEGDILAE 1724

Qy 1616 HFSAKDVVGDERYEAVRNLKKLVIRQQAADSHS--------MELGSASHSTT 1659
Db 1725 CINSAMPKG-KSHKPFR-VKKIMDQVQQASSTSSGANKNQVDTKKKKPTSPVKPMPQNTE 1782

Qy 1660 YNNLSRINNDGIV-------ELLHKHFDAALPA--SSAKRLGEMM-NNDPALKDII--- 1705
Db 1783 YRTRVRKNTDSKVNVNTEETFSDNKDSKKPSLQTNAKAFNEKLPNNEDRVRGTFALDSPH 1842

Qy 1706 --KQLQSTPF----------SSASVSMELKDGLREQTEKAILDGKVGREEVGVLFQDRNNLR 1755
Db 1843 HYTPIEGTPYCFSRNDSLSSLDFDDDVDLSREKAE--LRKGKESKDS---EAKVTCR 1895

Qy 1756 VKSVSVSQSVSKSE-GFNTPALLGTSNSAAMSMERNIGTINFKYGQD----- 1802
Db 1896 PEPNSSQQAASKSQASIKHPA--NRAQSKPVLQKQPTFPQSSKDGPDRGAATDEKLQNL 1952

Qy 1803 --QNTPRRFTLEGGIAQANPQVASALTDLKKE 1832
Db 1953 AIENTPVCFSRNSSL--------SSLSDIDQE 1976
```

RESULT 9
INVO_PONPY
ID INVO_PONPY   STANDARD;   PRT;   835 AA.
AC P14708;

```
Qy    4 KSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQPSTA 63
Db  242 RSSQSRHDAASHEAGRQHEGHGVAESNTAASSSGQSPAT----------RVDHETAS 288

Qy   64 ADGISAAHQQKKSFSLRGCLGTK--KFSRSAPQGQPGTTHSKGATLRDLLARDDGETQH 120
Db  289 VLSSSGTHSAPR--RLTSHLGTKVEMVSLLSMLG---THDKDDMSRTLLAM--- 335

Qy  121 EAAAPDAARLTRSGGVKRRNMDDMAGRPM--VKGGSGEDKVPTQQKRHQLNNFGQMRQT 177
Db  336 -SSSQDSCISMRQSGC----------LPLLIQLLHGNDKDSVLLGNSRG------- 373

Qy  178 MLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSE 237
Db  374 --SKEARARASAALHNIIHSQPDDKRGRREIR--------VLHLLEQIRAYCETCWE 420

Qy  238 FQQLHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSML 297
Db  421 WQEAHEQGMDQDKNPMPAPVEHQIC--------- 445

Qy  298 KGSGAGVTPLAVTLDKGKIQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSD--GSQHL 354
Db  446 ------PAVCVLMK--LSFDEEH--RHAMNELGGLQAIA 474

Qy  355 -LLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSH 413
Db  475 ELLQVDCEMYGL--TNDHYSVTLRRYAG-----------MALTNLTFGDVAN 513

Qy  414 NKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHS 473
Db  514 KATLCSMKG-CMRALV----------AQLKSESEDLQQVIASVLR-NLS-WRA--DVNS 557

Qy  474 QLSRQADGKLYAL----------KDNRTLQ-------NLS----DNKSSEKLVDKIKSYSV-- 513
Db  558 KKTLREVGSVKALMECALEVKKESTLKSVLSALWNLSAHCTENKADICAVDGALAFLVGT 617

Qy  514 ----DQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGLL-HGKSELEAQ 568
Db  618 LTYRSQTNTLAIIES--------GGGILRNVSSLIATNEDHRQI-LRENNCLQTLLQHLK----- 668

Qy  569 SVAISHGRLVVADSEGKLF--SAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHD 626
Db  669 -----SHSLTIVSNACGTIWNLSARNPKD--------QEALWD-MGAVSMLKNLIHS 711

Qy  627 DHGQL---NALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEI-L 682
Db  712 KHKMIAMGSAAALRNLMANRPAKYKDANIMSPGSSLP--------SLHVRKQKALEAEL 762

Qy  683 DMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQL-KKGL--DGAAYLLKDGEVKRLNIN 739
Db  763 DAQHLSET---------FDNIDNLSPKASHRSKQRHKQNLYGDYAFDANRHDDSRSDNFN 813

Qy  740 QSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRS 799
Db  814 TGNMTVLSPYLNTTVLPSSSSSRGSLDSSRS-EKDRSLERERGIGLSAYHPTTENAGTSS 872

Qy  800 FQIKPGTQQLERPAQTLSREGISSGELKDIHVDHKQNLYALTHEGEVFHQPREAWQNGAES 859
Db  873 ---KRGLQITTTAAQIAK---VMEEVSAIHTSQDDRSSASTTE---FHCVADDRSAARRS 923

Qy  860 SSWHKLA----LPQSESKLKSLDM----------SHEHKPIATFED--GSOHQLKAGGWH 903
Db  924 SASHTHSNTYNFTKSENSNRTCSMPYAKVEYKRSSNDSLNSVTSSDGYGKRGQMKP-SVE 982

Qy  904 AYAAPERGPLAVGTSGSGQTVFNRL------MQGVKGKVIPGSGLTVKLSAQ---TGGMTGA 955
Db  983 SYSEDDDESKFCSYGQYPADLAHKIHSANHMDDNDGELDTPINYSLKYSDEQLNSGRQSPS 1042

Qy  956 EGRKVSS-------------KFSERIRAYAFN---PTMSTPRPIKNAAYATQHGWQGREGLKPL 1003
Db 1043 QNERWARPKHVIEDEIKQNEQRQARSQNTSYPVYSENTDDKHLKFQPHFGQQ--ECVSP- 1099
```

```
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   INVOLUCRIN.
GN   IVL.
OS   Pongo pygmaeus (Orangutan).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX   NCBI_TaxID=9600;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90014143; PubMed=2796727;
RA   Djian P., Green H.;
RT   "The involucrin gene of the orangutan: generation of the late region
RT   as an evolutionary trend in the hominoids.";
RL   Mol. Biol. Evol. 6:469-477(1989).
 -!- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC       IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC       MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC       FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC   -!- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC       OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M25312; AAA36935.1; -.
DR   InterPro; IPR000354; -.
DR   InterPro; IPR002360; -.
DR   Pfam; PF00904; Involucrin; 1.
DR   PROSITE; PS00795; INVOLUCRIN; 1.
KW   Keratinocyte; Repeat.
SQ   SEQUENCE   835 AA;  97978 MW;  2EE73E73E1EA41D4 CRC64;
```

```
Query Match       2.1%;    Score 198.5;  DB 1;  Length 835;
Best Local Similarity  20.4%;   Pred. No. 0.026;
Matches  208;  Conservative 136;  Mismatches 400;  Indels 277;  Gaps 51;

Qy   185  PASANAGDRLQHSPPHIPGSHHEIKEE-PVGSTSKATTAHADRV-------EIAQEDD 234
               | :  | |  ||:|   ||   ||||    | :  || ||  ||
Db    24  PPPVNTQQEQMKQPTPLPPPCQKVPVELPVEVPSKQEEKHMTAVKGLPEQECEQQQQEPQ 83
Qy   235  DSEFQQLH--QQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLK 292
               :|||     || |    ||  | |        ||:  |
Db    84  EQELQQQHWEQHEEHQKAENPEQQLK------QEK-AQRDQQLNEQLEEEKLL 130
Qy   293  PQSMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQ 352
                |          | | |            ::|:|  | |
Db   131  DQR----------LDQELVKRDEQLGMKKEQLLE-----LPEQQEQHLKHLEQQEGQL 173
Qy   353  HLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSKGKISLGSGTQS 412
                |  |        :|             |:|  | |   :|: |    |
Db   174  ELPEQQEGQLKHLE---------QQEGQLKHLEQQEGQLEVP-EEQVGQLKYLEQQEG 221
Qy   413  HNKTMLSQPGE-AHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDT 471
               |: | |     |   ::|      | |  | ::|::|  | | |
Db   222  QLKHLDQQEGQLKHLDQQEGQLKH--LDQQEGQLKHLDQQEGQLKHLD-------QQEG 271
Qy   472  HSQLSRQADGKLYALKDNR-TLQNLSDNKSS---EKLVDKIKSYSVDQRGQVAILTDTP 526
               |  :    || | |  :      |   |:       ::| |    || | |  |
Db   272  QLELPEQQEGQLKHLEQQEGQLKHLEHEEGQLEVPEEQVGQLK-YLEQQEGQLKHLDQQE 330
Qy   527  GRHKMSIMPSLDASPESHISLSLHFADAHQGLL-----HGKSELEAQSVAI-------SH 574
               | :|        | |       | | ||          : :||
Db   331  GQLEL--------PEQQEGQLKHL-EQQEGQLKHLEHQKGQLEVPEEQVGQLKYLEQQE 380
Qy   575  GRLVVAD-SEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDHGQLNA 633
```

```
Db   381  GQLKHLDQQEGQL--ELPEQQEG-QLKHLEQQEGQL-KHLEHQEGQLEVPEEQVGQLKY 435
               |:|:| |::| ||          ::  |   |       :| |         |||
Qy   634  LVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDNLQG-LHHTNPEPHEI---LDMGHLG 688
Db   436  LEQQEGQLKH--                LDQQEGQLKHLDQQEKQLELPEQQVGQLK 474
                ||:|| ||                    : |  |: :
Qy   689  SLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYL-LKDGEVKRLN------IN 739
Db   475  HLEQQEGQLEV--------PEQVGQLK------YLEQQEGQLKHLDQQEGQLELPE 517
                |::  ||                  ::|:|:|
Qy   740  QSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDKAQAMAVIGVNKYLALTEKGDIRS 799
Db   518  QQEGQLKHLEQQEGQLKHLEHQ---EGQLEVPEEQ---VGQLKYLE-QQEGQLKH 565
                | |  |:||:| ||:|:|          :| |          | ||
Qy   800  FQIKPG-------TQQLERPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFHQPRE 851
Db   566  LDQQEGQLKHLDQQEKQLELPEQQV-------GQLKHLEQQEGQLEHLEQQEGQLEHLEHQ 619
                | |  ::| |: |: |:::|           |  |::|  |::| | | |:|
Qy   852  AWQNG-AESSSWHKLALPQSESKLSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPER 910
Db   620  EGQLGLPEQQVWQLKQLEKQEGQPKNLEE--------EEG---QLK------ 654
                | |  |::                             ::  |||
Qy   911  GPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRA 970
Db   655  ------HLVQ-QEGQLEQQEGQVEHLEEQVGQLKHLEEQEGQLKYLEQQQG 698
                     ::|   ||:||::| |:::|:|          :||  | | |::
Qy   971  YAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAP----Q 1025
Db   699  QLEVPEQQVGQP--------KHLEQEKQLE-LPEQQEGQLKHLEKQEAQLELPEQQVGQ 749
                ::|  | |:            :| |  |   ||::| |::||:|:||::| |
Qy  1026  P-DLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSE-F 1083
Db   750  PKHLEQQEKQL--EHPEQKDGQLK------HLEQQE------GQLKNLEQQKGQLEQPVF 795
                 |  |:|:          || ||           |       :|: :|::|:|  :|
Qy  1084  KPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPSAESKLQSMLGHFVSAGVDMSHQKGE 1143
Db   796  APAPGQV--------QDIQPAL-------PTKGE-----------VLLPVEQQQQKQE 827
                | | |:         ::| |        || |              | |::|:| |
Qy  1144  I 1144
Db   828  V 828
```

```
RESULT 10
P2P_LACPA
ID   P2P_LACPA     STANDARD;      PRT;   1902 AA.
AC   Q02470;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
DE   ASSOCIATED SERINE PROTEINASE) (LP151).
GN   PRTP.
OS   Lactobacillus paracasei.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=1597;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCDO 151;
RX   MEDLINE=92381481; PubMed=1512565;
RA   Holck A., Naes H.;
RT   "Cloning, sequencing and expression of the gene encoding the cell-
RT   envelope-associated proteinase from Lactobacillus paracasei subsp.
RT   paracasei NCDO 151.";
RL   J. Gen. Microbiol. 138:1353-1364(1992).
RN   [2]
RP   SEQUENCE OF 189-196.
RX   MEDLINE=92226694; PubMed=1564442;
RA   Naes H., Nissen-Meyer J.;
RT   "Purification and N-terminal amino acid sequence determination of the
RT   cell-wall-bound proteinase from Lactobacillus paracasei subsp.
```

```
RT   paracasei.";
RL   J. Gen. Microbiol. 138:313-318(1992).
CC   -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC       GROWTH OF THE BACTERIA ON MILK.
CC   -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC       SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC       E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC       PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC       ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC       INSULIN B-CHAIN.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC       SUBTILASE FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; M83946; AAA25248.1; -.
DR   PIR; B44858; B44858.
DR   HSSP; Q99405; 1MPT.
DR   MEROPS; S08.019; -.
DR   InterPro; IPR000209; -.
DR   InterPro; IPR001899; -.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF00082; Peptidase_S8; 3.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW   Hydrolase; Serine protease; Cell wall; Zymogen; Signal;
KW   Transmembrane.
FT   SIGNAL         1     33       POTENTIAL.
FT   PROPEP        34    187       POTENTIAL.
FT   CHAIN        188   1902       PI-TYPE PROTEINASE.
FT   DOMAIN       188   1876       EXTRACELLULAR (POTENTIAL)
FT   TRANSMEM    1877   1895       MEMBRANE ANCHOR (BY SIMILARITY).
FT   DOMAIN      1896   1902       CYTOPLASMIC (POTENTIAL).
FT   ACT_SITE     217    217       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE     281    281       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE     620    620       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DOMAIN      1867   1872       CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT                                 PROTEINS.
SQ   SEQUENCE   1902 AA;  200253 MW;  D8C9F38CEE5DA582 CRC64;
```

```
Db   235 VKLTKYDVEKFTDTAKHGRYFTSKVPYGFNYADNDTITDDTVDEQHGMHVAGIIGANGT 294
Qy   253 PPQPPK--LGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAVT 310
Db   295 GDDPFKSVVGVA-------PEAQLLAMKVFTNSDT-------SATTGSATLVSAIE-- 336
Qy   311 LDKGKQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTAT 370
Db   337 -DSAKIGA-------DVLNMSLGSDS----------GNQTL-------------EDP 362
Qy   371 SYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNK--------TMLSQPG 422
Db   363 EIAAVQNAN-------ESGTAAVISAGNSG--TSGSATQGVNKDYYGLQDNEMVGTPG 411
Qy   423 EAHRSLLTGIWQHPAGAARPQ--GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480
Db   412 TSRGATTV-------ASAENTDVISQAVTITDGKDLQLGPE--TIQLSSNDFTGSFDQK-- 461
Qy   481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540
Db   462 -KFYVVKD-------ASGDLSKGAAAD---YTADAKGKIAIVKRG-------- 495
Qy   541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEG---------- 584
Db   496 ------ELNFADK-------QKYAQAGAAGLLIVNNDGTATPLTSIRLTTTFPT 537
Qy   585 ------------KL-FSAAIPKQGDNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNA 633
Db   538 FGLSSKTGQKLVDWVTAHPDDSLGVKIALTLLPNQKYE-----DKMSDF--TSYGPVSN 590
Qy   634 LVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHGLGSLALQ 693
Db   591 L---SFKPDITAPGGN-------IWSTQNNNGYTNMSGTSMASPFIAGSQALL 633
Qy   694 EGKLH-----YFDQLTKGWTGAESDCKQLKKGLDGAAY--LLKDGEVKRLNINQSTSSIK 746
Db   634 KQALNNKNNPFY-------ADYKQLK---GTALTDFLKTVE--MNTAQPINDIN 675
Qy   747 HGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKG--DIRSFQI 802
Db   676 Y--NNVIVSPR-------RQGAGLVDVKAAIDALEKNPSTVVAENGYPAVELKDFTS 723
Qy   803 KPGTOQL-------ERPAQTLSREGISGELKDIHVDHKQNLYALTHEG 843
Db   724 TDKTFKLTFTNRTTHELTYQMDSNTDTNAVYTSATDPNSGVLYDKKIDG----AAIKAG 778
Qy   844 EVFHQPREAWQNGAESSWHKLALPQSESKLKSLDMSHEHKPIATFE--DGSQHQL--- 897
Db   779 SDITVPA------GKTAQIEFTLSLP-------KSFDQQQFVEGFLNFKGSDGSRLNLPYMG 827
Qy   898 ----------------------------------------WHAYAAPERGPLAVGTSGSQTVF 924
Db   828 FFGDWNDGKIVDSLNGITYSPAGGNYGTVPLLTNKNTGHQY---YGGMVTDADGKQTVD 883
Qy   925 NRLMQGVKGKVIPGSGLTVK-------LSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMS 978
Db   884 DQAIAFSSDKNALYNDISMQYYLLRNISNVQVDILDGGGNKVTTLSSSTNQTKTYDAHS 943
Qy   979 TPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLD-------- 1015
Db   944 ----QKYIYYNAPAWDG---TYIDQRDGNIKTADDGSYTYRISGVPEGGDKRQVFDV 993
Qy  1016 ------AHNVRHNAPQPDLQSKLET------------LDLGEHGAELLNDMKR- 1050
Db   994 PFKLDSKAPTVRHVA---LSAKTENGTKQYYLTAEAKDDLSGLDATKSVKTAINEVTNL 1049
Qy  1051 ---FRD--ELEQSATRSVTVLGQHQGVLKSNGEINSE--FKPSPGKALVQSFNVNRSGQ- 1102
Db  1050 DATFTDAGTTADGYTKIETPLSDEQAQALGNGDNSAELYTDNASNATNQDASVQKPGST 1109
Qy  1103 --DL--SKSLQQAVHATPPSAESKLQ----SMLGHFVSAGVDMSHQKGEIPLGRQDP 1152
Db  1110 SFDLIVNGGGIPDKISSTTTGYEANTQGGGTYTFSGTY-PAAVDGTYTDAQ--GKKHDL 1165
```

```
Query Match          2.1%;    Score 196.5;  DB 1;  Length 1902;
Best Local Similarity  18.9%;   Pred. No. 0.11;
Matches  397;  Conservative 248;  Mismatches 747;  Indels 707;  Gaps 102;

Qy     2 ELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQPS 61
           ::  ||   ||| ||||     |
Db    40 KVSSLANTVKAA--TAKQ------AATDTTAATTNQAIATQLAAKGIDYNKLNKVQQQD 90
Qy    62 TAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQHE 121
           || | |  :  :  ::
Db    91 TYVDVIVQ-------MSAAPASENGTLRTDYSS------------TAEIQQE 123
Qy   122 AAAPDAARLTRSGGVKRRNMDDMAGRP--MVKGGSGEDKVPTQQKRHQLNNFGQMRQTM 178
           |||  :|  :||  :  |      ::
Db   124 TNKVIAQASVKAAVEQVT-QOTAGESGYGVVNGFSTKVRVVDIPK--LKQIAGVKTVT 179
Qy   179 LSKMAHPASANAGDRLQHSPPHIPGSHHEIKEPVGS----------215
           |||: | ||  | |: |
Db   180 LAKVYYPTDKA-----NSMANVQAVWSNYKYKGEGTVVSVIDTGIDPTHKDMRLSDDKD 234
Qy   216 ---------------TSKATTA-HADRVEIAQEDDDSEFQQLHQQRLAREREN 252
                          |||  : |   | :|  ::   |  : | : | |
```

```
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VIII.
RT   78 new cDNA clones from brain which code for large proteins in
RT   vitro.";
RL   DNA Res. 4:307-313(1997).
CC   -!- FUNCTION: CENTROSOMAL PROTEIN.
CC   -!- SIMILARITY: STRONG, TO MOUSE PERICENTIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U52962; AAD10838.1; -.
DR   EMBL; AB007862; BAA23698.1; -.
DR   HSSP; P02649; 1LPE.
KW   Coiled coil.
FT   DOMAIN      258    553   COILED COIL (POTENTIAL).
FT   DOMAIN      675    835   COILED COIL (POTENTIAL).
FT   DOMAIN      999   1135   COILED COIL (POTENTIAL).
FT   DOMAIN     1288   1938   COILED COIL (POTENTIAL).
FT   DOMAIN     2053   2071   COILED COIL (POTENTIAL).
FT   DOMAIN     2525   3075   COILED COIL (POTENTIAL).
FT   CONFLICT   1523   1523   H -> Q (IN REF. 2).
FT   CONFLICT   2177   2177   R -> M (IN REF. 2).
FT   CONFLICT   2538   2538   T -> A (IN REF. 2).
FT   CONFLICT   2828   2906   MISSING (IN REF. 2).
FT   CONFLICT   3012   3012   K -> E (IN REF. 2).
FT   CONFLICT   3125   3125   S -> Y (IN REF. 2).
FT   CONFLICT   3289   3289   G -> E (IN REF. 2).
FT   CONFLICT   3309   3321   YYQILQRNPATR -> VLPDSTSKKSCHPMIKQ (IN
FT                            REF. 2).
SQ   SEQUENCE   3321 AA;  376327 MW;  0DB3D287BF7D8762 CRC64;

Query Match
Best Local Similarity   2.1%;   Score 195;   DB 1;   Length 3321;
Matches  290;   Conservative  256;   Mismatches 588;   Indels 570;   Gaps  72;
Similarity  17.0%;   Pred. No. 0.32;

Qy  377 NSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHP 436
                                                             ||       |  ||
Db  110 NDHPPE------------QCGMFTVS-----------------------DHPPEQHGMFTVG-DHP 139

Qy  437 AGAARPQGESIRLHDDKIHILHP-ELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQN 494
              ||                                                      |
Db  140 -----PEQRGMFTVSD--------HPPEQHGMFTVSDHP-----------PEQRGMFT 173

Qy  495 LSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHIS--- 546
        ::|::   |::|   :::|                |   |   | : |
Db  174 ISDHQPEQRGMFTVSDHTPEQRG-IFTISDHPAEQRGMFTKECEQECELAITDLESGRED 232

Qy  547 -LSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAM 605
          ||| | |     |||||   |         :::                    ||
Db  233 EAGLHQSQAVHGL---ELEALRLSLSN---MHTAQLELTQANLQKEKETALTELREM 283

Qy  606 PQHALDEHFGHDHQISGFFHDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVI 665
               |                |||   :||::|
Db  284 --------------LNSRRAQELALLQS-RQQHELEL----------------- 305

Qy  666 DNQLGLHHTNPEPHEILDMGHLGS------LALQEGKLHYFDQLTKGWTGAESDC-KQLK 718
                                         |||  ||||| ||::
Db  306 --LREQHAREKEEVVLRCGQEAAELKEKLQSEMEKNAQIVKTLKEDWESEKDLCLENLR 362

Qy  719 KGLDGAAYLLKDGEVKRLNLNQSTSSIKHGTE-----NVFSLPHVRNKPEPGDALQGLNK 773
        |::|                                     :|
Db  363 KEL----------SAKHQSEMEDLQNQFQKELAEQRAELEKIFQDKNQ 400

Qy  774 DDKAQAMAVIGVNKYLALTEK-GDIRS------------FQIKPGTQQLERPAQTLSREGI 821
        ::|                                    :|:
Db  401 AERALRNL--ESHHQAAIEKLREDLQSEHGRCLEDLEFKFKFKESEKE----KQLELENL 452
```

```
Qy 1153 ND------KTALTKSRLILDTVTIGELHELADAKAKLVSDHKPDADQIKQLRQQPFDTLREKR 1207
                 | :  :  ::  ||||                                  ||  |  ||
Db 1166 NTTYDAATNSFTASMAVTNADYAAQVDLYADKAH------------TQLLKHFDT--;KV 1210

Qy 1208 YESNPVKHYTDMGFTH-NKALEANYDA------------VKAFINAFKKEHH------ 1246
Db 1211 RLTAPT--FTDLKFNNGSDQTSEATIKVTGTVSSDTKTVNVGDTVAALDAQHHFSVDVPV 1268

Qy 1247 --GVNLTTRTVLESQGSAELAKK--------------LKNTLLSLDSGESMSFSRSYGGGVSTV 1294
             |  :  | : :: | :                 :  | : ||  | :
Db 1269 NYGDNTIKVTATDEDGNTTEQKTITSSYDPDVLKNA-VKFGANEFNATSAK 1325

Qy 1295 F------VPTLSKKVPVPV----IPGAGITLDRAYNLSFSRTSGGLNVS-FGRDGGVSGN 1343
           :::: | :  :                  ||   :|  :::
Db 1326 FYDPKTGIATITGKVKHPTTLQVDGKQISIKNDLTFSFTLDGTLGQKPFG------ 1377

Qy 1344 IMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQCTLQNSLKFKLT 1403
Db 1378 --VVVGD------TTQNKTFQEALTFILDA--VAPTLSLD--SSTDAPVYTNDPNFQIT 1424

Qy 1404 EDELPGFIHGLTHGTLTP-AELLQKGI-EHQMKQGSKLTFSVDTSANLDLRAGINLNE- 1459
        :::  :        ::    :::  :   :: ::::  :      | ::
Db 1425 ------GTATDNAQYLSLAINGSHVASQYADININSGKPGHMAIDQPVKLLEG 1471

Qy 1460 ------DGSKPNGVTARVSRAVSAGLSASANLAAGSRERSTT------SGQFGSTTS 1500
               |:|             ||    :|:   |||
Db 1472 KNVLTVAVTDSENNTTTKKITVYYEPKKTLAAPTVTPSTTEPAKTVTLTANAAATGETVQ 1531

Qy 1501 -ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIP----PA--FTSTNVSAALALD 1553
         |:|:       |  ||  |          |::         ||  |:|
Db 1532 YSADGGKTYQDVPAAGVTVTA------NGTFKFKSTDLYGNESPAVDYVVTNIKA----- 1580

Qy 1554 NRTSQSISLELKRAEPVTSNDI--SELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLH 1611
          : : ::   :  ::       ::| |||  |    | |  ||||
Db 1581 -----DDPAQLQTAKQALTNLIASAKTLSASGKY--DDATTTALAA------ATQKAQTA 1627

Qy 1612 ILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGI 1671
         | ::|   :             |:|                     :::     ||
Db 1628 LDQTDASVDSLTG------ANRDLQ----------------TAINQLA------- 1653

Qy 1672 VELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTE 1731
        ::   ::    :              ||   ::| |:|
Db 1654 ------AKLPADKKTSL-----LNQLQSV--KAALGTDLGNQTDPSTG 1688

Qy 1732 K---AILDGKVGREEVGVLFQDRNNLRVKSVSVSQSVSKSEGFN--TPALLGTSNSAA 1785
        |        |  :: |           ::: ::: | ::
Db 1689 KTFTAALDDLVAQAQAGTQTADQLQASLAKVLDAVLAKLAEGIKAATPA-EVGNAKDAA 1746
```

```
RESULT 11
KEND_HUMAN
ID   KEND_HUMAN    STANDARD;    PRT;   3321 AA.
AC   O95613; O43152;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   KENDRIN.
GN   KIAA0402.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Li Q., Joshi H.C.;
RT   "Kendrin, an integral component of centrosome.";
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1512-3321 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98116655; PubMed=9455477;
RA   Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
```

```
Qy  822  SGELKDIHVDHKQNLYALTHEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSH  881
Db  453  QASYEDLKAQSQEIRRL------WSQLDSARTSRQELSELHEQLLART-SRVEDLEQLK    505

Qy  882  EHKPIATPEDGSOHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVGKVIPGSGL   941
Db  506  Q----REKTQHESELEQLRIYFEKKLRDAEKTYQEDLTLLQQRLQGAREDALLDS-V     557

Qy  942  TVKLSAQTGGMTGAEGRK------VSSKFSERIRAYAFN------PTMSTPRPIKNA     986
Db  558  EVGLSCVGLEEKPEKGRKDHVDELEPERHKESLPRFQAELEESHRHQLEALESPLCIQHE  617

Qy  987  AY------QGREGLKPLYEMQGALIKQLDA-------H                      1017
Db  618  GHVSDRCCVETSALGHEWRLEPSEGHSQELPWVHLQGVQDGDLEADTERAARVLGLETEH  677

Qy 1018  NVRHNAPQPDLQSKLETLD------LGEHGAELLN----                     1046
Db  678  KVQLSLLQTELKEIELLKIENRNLFEKLQHETRLKDDLEKVKHNLIEDHQKELNNAKQK   737

Qy 1047  ---------DMKRFRDELEQSATRSVTVL-------GQHQGVLKSNGEINSEFK       1084
Db  738  TELMKQEFQRKETDWKVMKEELQREAEEKTLMLELELREKAESEKQTIINKLELREAEMR  797

Qy 1085  PSPGKALVQSFNVNRSGQDLSKSLQQ--------AVHAT-----PPSAES-------    1121
Db  798  QLQDQQAAQILDLERSLTEQQARLQQLEQDLTSDDALHCSQCGREPPTAQDGELAALHVK  857

Qy 1122  ---KLQSML--GHFVSAGVDMS--------HQK------GEIPLGRQRDPN          1153
Db  858  EDCALQLMLARSRFLEERKEIAKFSAEQDAFLQLLQERHQQQLLSVTAELEARHQAALG   917

Qy 1154  DKTALTKSR---LILDTVTIGELHELADKAKLVSDHKPDADQIKQ-LRQQFDTLREKRYE  1209
Db  918  ELTASLESKQGALLAARVAELQTKHAADIGALETRHLSTLDSLESCYLSEFQTIREE---   974

Qy 1210  SNPVKHYTDMGFTHNKALE---ANYD----------AVKAFINAFKKEHHGVNLTTRT   1254
Db  975  ---HRQALELLRADFEEQLWKKDSLHQTILTQELEKLKRKHHKGELQSVRD         1021

Qy 1255  VLESQGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGIT 1314
Db 1022  HLRTEASTELAGTVAHELQGVHQGE------------------                 1046

Qy 1315  LDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKH 1374
Db 1047  ------FGSE--------------KKT------------ALH                  1056

Qy 1375  KISPDLRIGAAVSGTLGQSLKFKLTE-----DELPGFIHGLTH--GTLTPAELLQ     1426
Db 1057  EKEETLRLQSAQAQPFHQEEKESLSLQLQKKNHQVQQLKDQVLSLSHEIEECRSELEVLQ 1116

Qy 1427  KGIEHQMKQGSKLTFSVDTSANLDLRAGINL--NEDGSKPNGVTARVSAGLSASANLAAG 1484
Db 1117  --------LKADVNLSHSERGALQDALRRLL----------                   1155

Qy 1485  SRERSTTSGQFGSTTSASNNRPTFLNGVGAGANL---TAALGVAHSSTHEGKPVGIFPAF 1541
Db 1156  ------GLFGETLRAA--VTLRSRIGERVGLCLDDAGAGLALSTA--------PAL     1195

Qy 1542  TSTNVSAAL-ALDNRTSQ------SISLELKRAEPVTSNDISELTSTLGKHFK----    1587
Db 1196  EETWSDVALPELDRTLSECAEMSSVAELSSHMRESFLMSPESVRECEQPIRRVFQSLSLA 1255

Qy 1588  -DSATTKMLAALKELDDAKPAEQLHI-LQQHFSAKDVVGDERYEAVRNLKKLVIRQQAAD 1645
Db 1256  VDGLMEMALDSSRQLEEAR---QLHSRPEKEFSFKN----EETAQVVRKHQELLECLTEES 1309

Qy 1646  SHSMELGSASHST--TYNNLSRINNDGIVELLHKHFDA---ALPASSAKR-LGEMMNND  1698
Db 1310  AAKAELALELHKTQGTLEGF-KVETADLKEVLAGKEDSEHRLVLELESLRRQLQQAAQEQ 1368

Qy 1699  PALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVGREEVGVLFQDR----NNL 1754
Db 1369  AALREECTRLWSRGEATATDAEAREAALRKEVEDLTKEQSETRKQAE---KDRSALLSQM 1425

Qy 1755  RVKSVSYSQSVSKSEGFNTPALLGTSNSAAMSERNIGTI----NFKYGQDQNTPRR-    1808
Db 1426  KILESELEEQLSQHRGCAKQA-------EAVTALEQQVASLDKHLRNQRQFMDEQAAERE 1478

Qy 1809  -----------FTLEGGIAQ-ANPQ                                   1821
Db 1479  HEREEFQQEIQRLEGQLRQAAKPQ                                    1502
```

RESULT 12

MAP2_RAT

```
ID   MAP2_RAT        STANDARD;        PRT;     1861 AA.
AC   P15146;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-OCT-1994 (Rel. 30, Last annotation update)
DE   MICROTUBULE-ASSOCIATED PROTEIN 2 (MAP2B) [CONTAINS: MAP2C].
GN   MAP2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.
RC   STRAIN=WISTAR; TISSUE=Brain;
RX   MEDLINE=90251471; PubMed=2339070;
RA   Kindler S., Schwanke B., Schulz B., Garner C.C.;
RT   "Complete cDNA sequence encoding rat high and low molecular weight
RT   MAP2.";
RL   Nucleic Acids Res. 18:2822-2822(1990).
RN   [2]
RP   SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.
RC   STRAIN=WISTAR; TISSUE=Brain;
RX   MEDLINE=91060576; PubMed=2174050;
RA   Kindler S., Schulz B., Goedert M., Garner C.C.;
RT   "Molecular structure of microtubule-associated protein 2b and 2c from
RT   rat brain.";
RL   J. Biol. Chem. 265:19679-19684(1990).
RN   [3]
RP   SEQUENCE OF 1-151; 1515-1694 AND 1726-1861 FROM N.A.
RX   MEDLINE=90221819; PubMed=2326166;
RA   Doll T., Papandrikopoulou A., Matus A.;
RT   "Nucleotide and amino acid sequences of embryonic rat MAP2c.";
RL   Nucleic Acids Res. 18:361-361(1990).
RN   [4]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=89365159; PubMed=2770869;
RA   Papandrikopoulou A., Doll T., Tucker R.P., Garner C.C., Matus A.;
RT   "Embryonic MAP2 lacks the cross-linking sidearm sequences and
RT   dendritic targeting signal of adult MAP2.";
RL   Nature 340:650-652(1989).
RN   [5]
RP   SEQUENCE OF 1695-1725 FROM N.A.
RX   MEDLINE=94110302; PubMed=8282767;
RA   Doll T., Meichsner M., Riederer B.M., Honegger P., Matus A.;
RT   "An isoform of microtubule-associated protein 2 (MAP2) containing
RT   four repeats of the tubulin-binding motif.";
RL   J. Cell Sci. 106:633-640(1993).
CC   -!- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY
CC       STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO
CC       SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.
CC   -!- ALTERNATIVE PRODUCTS: VARIOUS FORMS OF MAP2 ARE PRODUCED BY
CC       ALTERNATIVE SPLICING OF THE SAME GENE. MAP2C, THE LOW MOLECULAR
CC       FORM OF MAP2, LACKS THE CENTRAL DOMAIN OF MAP2A/B.
CC   -!- DEVELOPMENTAL STAGE: MAP2C IS EXPRESSED DURING EMBRYONIC BRAIN
CC       DEVELOPMENT AND UNTIL POSTANATAL DAY 10. MAP2B IS EXPRESSED
CC       THROUGHOUT BRAIN DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 3 OR 4 TAU/MAP REPEATS.
```

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).

DR   EMBL; X51842; CAA36135.1; -.
DR   EMBL; X17682; CAA35667.1; -.
DR   EMBL; X71487; CAA50588.1; -.
DR   PIR; S07887; S07887.
DR   PIR; S10003; S10003.
DR   PIR; A37981; A37981.
DR   InterPro; IPR001084; -.
DR   Pfam; PF00418; tubulin-binding; 4.
DR   PROSITE; PS00229; TAU_MAP_1; 3.
KW   Microtubules; Repeat; Alternative splicing; Calmodulin-binding.
FT   DOMAIN   1454   1474   CALMODULIN-BINDING (POTENTIAL).
FT   REPEAT   1664   1694   TAU/MAP MOTIF.
FT   REPEAT   1695   1725   TAU/MAP MOTIF.
FT   REPEAT   1726   1756   TAU/MAP MOTIF.
FT   REPEAT   1757   1788   TAU/MAP MOTIF.
FT   VARSPLIC   152   1514   MISSING (IN ISOFORM MAP2C).
FT   VARSPLIC   1695   1725   MISSING (IN ISOFORM WITH 3 TAU/MAP
FT   REPEATS).
SQ   SEQUENCE   1861 AA; 202409 MW; 42DCF116D21EF54E CRC64;

Query Match   2.1%; Score 194; DB 1; Length 1861;
Best Local Similarity 18.9%; Pred. No. 0.15;
Matches 407; Conservative 274; Mismatches 819; Indels 652; Gaps 98;

Qy   44 LAAEGKNRGKMPRIHQPSTAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSK 103
         ::|:|| | |:|| ||||| : :| ||| ||| : | : ||
Db   1 MADERKDEGKAP--HWTSASLTE-AAAHPSPEMKDQGSSG-EGLSRSA-NGFPYREEEE 55

Qy   104 GATLRDLLARDDGETQHEAAAAPDAARLTRSGGVK--------RRNMDDMAGR------- 147
         |: ||:|::| | |||: :|: : |:   :||
Db   56 GAF---------GEHGSQGTYSD---TKENGINGELTSADRETAEEVSARIVQVVTAEA 102

Qy   148 -PMVKGGSGGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHH 206
         |:| : || ||: || : | :| : :|: |:|||::|:
Db   103 VAVLKG----------EQEKEAQHKD-----QPAALPLAAEETVNLPPSPPPSPASEQ 145

Qy   207 EIK-EEPVGSTSKATTAHADRV-EIAQEDDSEFQQLHQQRLARERENP--PQPP----- 257
         |:| ||::| | || :|| : :| |: || ||: :|:   :||
Db   146 TAALEEDLLTASKMEFPEQQKLPSSFAEPLDKEETEFKMQSKPGEDFEHAALVPQPDTSK 205

Qy   258 ---------KLGVATPISARFQPKLTAVAESVLEGTDTTQS------PLKPQSMLK 298
         ||| |:|| | |:|:   :| | | |: :|
Db   206 TPQDKKPDQDMEGEKSPASPFAQTFGTNLEDI-KQITEPSITVPSIGLSAEPLAPKDQKD 264

Qy   299 ---------GSGAGVT--PLAVTLDKGKLQLAP------DNP-PALNTLLKQTL 334
         | | || ||| :|:   : | |:|
Db   265 WFIEMPVESKKDEWGLAAPISPGPLTPMREKDVLEDIPRWEGKQFDSPMPSPFHGGSFTL 324

Qy   335 GKDTQHYLAHHASSDGSQ-------HLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKG---- 385
         ||: :|| || | :| : || |:|| || :   :| :| |
Db   325 PLDT---VKDERVTEGSQPFAVFFQSDDKMSLQDTSGGSATSKESSKDEEPQKDKADKVA 381

Qy   386 --KLAQAGT----GSVSVDGKSGKISLG--SGTQSHNKTMLSQPGEAHRSLLTGIWQHP 436
         ::|: ||| | | : |||:|||:| ::| |
Db   382 DVPVSEATTVLGDVHSPAVEGFVGENISGEEKGTTDQEKKETSTPSVQEPTLT------ 434

Qy   437 AGAARPQ----------------GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQ 478
         | |: |||:||:| ||| ||: :: | |:
Db   435 --ETEPQTKLEETSKVSIEETVAKEEESLKLKDDK-------AGVIQTSEQSFSKEDQK 485

Qy   479 A-DGKLYALKDN----RTLQNLSDNKSSEKLVDKIKS--YSVDQRGQVAIL--TDTPGRH 529
         : ||| |||::   : :| |: :|:::: ||:|| : : | :| |:| |: :|
Db   486 GQEQTIEALKQDSFPISLEQAVTDAAMATKTLEKVTSEPEAVSEKREIQGLFEEDIADKS 545

```
Db   1513  TTATSGE---SAQAPSAFKQAKDV---TDGIT---------------------KSPEKRSSL 1548
                 ||  ||     :  :|  |       :|  :
Qy   1443  VDTSANLDRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTSAS 1502
                 :| |  ||  :    :  |  ::  |: |||::|||
Db   1549  PRPSSILPPRRCVSGDREEN----------SFSLNSSISARRTTRSEPIRRAGKSG 1595
                 |  ||  | | ||   :    :  ::
Qy   1503  NNRPFTFLNGVGAGANLTAALGVAHSSTHEGKP-VGIFPAFTSTNVSAALALDNRTSQSIS 1561
                 :  :    ||    :  |  | :      :  ||
Db   1596  TSTPT---TPGSTAITPGTPSYSSRTPGTPSYSYSSRTPGTPKSGILVPSEK------K 1646
                 :  |     |||   |  |:  |   |  ::
Qy   1562  LELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAKD 1621
                 : |     | | |      :  |:   :|: :  |
Db   1647  VAIIRTPP-----------------KSPATPKQLRLINQ---PLPDLKNVKSKIGSTD 1684
                 :  | |                 ||    : |
Qy   1622  VVGDERYE---AVRNLKKLVIRQQAADSHSMELGSA-SHSTTYNNLSRINNDGIVELLH 1676
                 ::    :|  : |||    |::        ::||::|
Db   1685  NI---KYQPKGGQVRILNK------------KMDFSKVQSRCGSKDNIKHSAGGNVQIVT 1730
                 |:   ||||||| |||                 :|  ||
Qy   1677  KHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILD 1736
                 |      |    |
Db   1731  KKID----------LSHVTSKCCGSLKNIRHRPGGGRVKIESVKLDFKE--KAQAKVGSLD 1778
                                :  | :      |  ::| :|     :
Qy   1737  -------GKVGREEVGVLFQDRNNLRVK--SVSVSQSVKSEGFNTPALLGTSNSAAM 1786
                                |  | |  |:|| :|:|  |  ||  ||:|
Db   1779  NAHHVPGGGNVKIDSQKLNFREHAKARVDHGAEIITQSPSRS-SVASPRRLSNVSSS--- 1834
                 :::|||| |:||    ||:|    | ::||:  : | :|||||::|
Qy   1787  SMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASALTD---LKKEGL 1834
                 :|:|                               :  ||:: | |:||
Db   1835  -----GSINL------------LE-----SPQLATLAEDVTAALAKQGL 1861
```

```
RESULT   13
PGCV_MOUSE
ID   PGCV_MOUSE     STANDARD;      PRT;    3358 AA.
AC   Q62059; Q62058;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   VERSICAN CORE PROTEIN PRECURSOR (LARGE FIBROBLAST PROTEOGLYCAN)
DE   (CHONDROITIN SULFATE PROTEOGLYCAN CORE PROTEIN 2) (PG-M).
GN   CSPG2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (VARIANTS V0; V1 AND V2).
RC   STRAIN=C57BL/6, AND SWISS WEBSTER; TISSUE=Brain;
RX   MEDLINE=95122551; PubMed=7822336;
RA   Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;
RT   "Multiple forms of mouse PG-M, a large chondroitin sulfate
RT   proteoglycan generated by alternative splicing.";
RL   J. Biol. Chem. 270:958-965(1995).
RN   [2]
RP   SEQUENCE OF 1-348 AND 3053-3358 FROM N.A. (VARIANT V3).
RC   STRAIN=C57BL/6;
RX   MEDLINE=95181355; PubMed=7876137;
RA   Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT   "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT   without a chondroitin sulfate attachment in region in mouse and human
RT   tissues.";
RL   J. Biol. Chem. 270:3914-3918(1995).
CC   -!- FUNCTION: MAY PLAY A ROLE IN INTERCELLULAR SIGNALING AND IN
CC       CONNECTING CELLS WITH THE EXTRACELLULAR MATRIX. MAY TAKE PART IN
CC       THE REGULATION OF CELL MOTILITY, GROWTH AND DIFFERENTIATION. BINDS
CC       HYALURONIC ACID.
CC   -!- SUBCELLULAR LOCATION: SECRETED; EXTRACELLULAR MATRIX.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS; V0 (SHOWN HERE), V1,
CC       V2 AND V3; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: V2 IS FOUND ONLY IN BRAIN.
CC   -!- DEVELOPMENTAL STAGE: DISAPPEARS AFTER THE CARTILAGE DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) REPEAT.
CC   -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
```

```
DR   EMBL; D16263; BAA03796.1; -.
DR   EMBL; D28599; -; NOT_ANNOTATED_CDS.
DR   EMBL; D32040; BAA06802.1; -.
DR   HSSP; P00740; 1IXA.
DR   MGD; MGI:102889; Cspg2.
DR   InterPro; IPR000152; -.
DR   InterPro; IPR000436; -.
DR   InterPro; IPR000538; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001304; -.
DR   InterPro; IPR001438; -.
DR   InterPro; IPR001881; -.
DR   InterPro; IPR003006; -.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00193; Xlink; 2.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PROSITE; PR00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS01241; LINK; 2.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW   Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW   Hyaluronic acid; Alternative splicing.
FT   SIGNAL         1     20       POTENTIAL.
FT   CHAIN         21   3358       VERSICAN CORE PROTEIN.
FT   DOMAIN        37    137       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       167    244       LINK 1.
FT   DOMAIN       265    334       LINK 1.
FT   DOMAIN       348   1308       GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
FT                                 DOMAIN).
FT   DOMAIN      1309   3052       GAG-BETA.
FT   DOMAIN      3052   3088       EGF-LIKE 1.
FT   DOMAIN      3090   3126       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      3129   3255       C-TYPE LECTIN.
FT   DOMAIN      3258   3316       SUSHI.
FT   DISULFID      44    130       BY SIMILARITY.
FT   DISULFID     172    243       BY SIMILARITY.
FT   DISULFID     196    217       BY SIMILARITY.
FT   DISULFID     270    333       BY SIMILARITY.
FT   DISULFID     294    315       BY SIMILARITY.
FT   DISULFID    3056   3067       BY SIMILARITY.
FT   DISULFID    3061   3076       BY SIMILARITY.
FT   DISULFID    3078   3087       BY SIMILARITY.
FT   DISULFID    3094   3105       BY SIMILARITY.
FT   DISULFID    3099   3114       BY SIMILARITY.
FT   DISULFID    3116   3125       BY SIMILARITY.
FT   DISULFID    3132   3143       BY SIMILARITY.
FT   DISULFID    3160   3252       BY SIMILARITY.
FT   DISULFID    3228   3244       BY SIMILARITY.
FT   DISULFID    3259   3302       BY SIMILARITY.
FT   DISULFID    3288   3315       BY SIMILARITY.
FT   CARBOHYD      57     57       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     330    330       N-LINKED (GLCNAC...) (POTENTIAL).
```

```
FT   CARBOHYD      351    351       N-LINKED (GLCNAC...   .)   (POTENTIAL).
FT   CARBOHYD      441    441       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD      807    807       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD      914    914       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD      951    951       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD     1305   1305       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD     1372   1372       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD     1679   1679       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD     2054   2054       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD     2244   2244       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD     2362   2362       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD     2627   2627       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD     3030   3030       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD     3332   3332       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   CARBOHYD     3342   3342       N-LINKED (GLCNAC...    )   (POTENTIAL).
FT   VARSPLIC      349   1308       MISSING (IN ISOFORM V1).
FT   VARSPLIC     1309   3052       MISSING (IN ISOFORM V2).
FT   VARSPLIC      349   3052       MISSING (IN ISOFORM V3).
FT   CONFLICT      348    348       P -> R (IN REF. 2).
SQ   SEQUENCE     3358 AA;  366938 MW;  071B80026BC0762D CRC64;

Query Match             2.1%;
Best Local Similarity  18.1%;   Score 194;   DB 1;   Length 3358;
Matches   388;  Conservative  287;  Mismatches  791;  Indels  680;  Gaps  104;

Qy    25 GVALQQGSSSSS---PQNAAASLAAEGKNRGK-----------MPRIHQPSTAAD-GIS   68
            : :                : :||: |                 |||  || :||:
Db   594 QGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQ----------  642

Qy   724 GMPLSTSSSESSVERTESVSPALTIE-KLTGKPTEARDVEEMTLTRLETDVTKSDKDVT  782
              :   :  : || :   : ||  |  |||  | ::|| ::|| :     |  | ||
Db  1397 DPEAAEAARGQYESVAPSQNFPDSSAT-----DTHQFILAETESSTTMQFKKSKEGTELL 1451

Qy   643 ----------HACPLGNDH----------QFHPG---W--NLTDALV-IDNQLGLHHTNPE  677
                      : |    |           | | |   |    |  |  | :| |: | |
Db  1452 EITWKPETYPETPDHVSSGEPDVFPTLSSHDGKTTRWSESITESSPNLENPV--HKQPK 1508

Qy   678 PHEILDMGHLGSLALQEGKLH-YFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRL  736
              | | |   ::| :|  |    :: | || |   | :|   | |:   :  | | :|:
Db  1509 PVPLFPEESSGEGAIEQASQETILSRATEVALGKETD----------------------  1546

Qy   737 NINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGD  796
              ||  |  |   :|  :| |        |  ::| |:  |  | :   :   |::|  |
Db  1547 SPTLSTSSILSSSVSVNVL-----EEEP-LTLGISQTDESMST----IESWVEIT---- 1592

Qy   797 IRSFQIKPGTQQLERPAQTLS---------REGISGELKDIHVDHKQNLYALTHEGEVFH  847
            | :| || :| :|             ||||  ||  : | :|| |  : ::|  || :
Db  1593 ---------PSQTVKFSESSSAPIIEG-SGEVEE----------NKNKIFN 1623

Qy   848 QPREAWQNGAESSSWHKLALPQSE-SKLKSLDMS-----HEHKP--IATFEDGSQHQ    896
            ||                 |||    |||:||         | |     |:
Db  1624 MVTD------------------LPQRDPTDTLSPLDMSKIMITNHIYIPATIAPLDS---- 1663

Qy   897 LKAGGWHAYAAPERGPLAV--GTSGSQTVFNRLMQGVGKGKVIPGSGLTVKLSAQTGGMTG  954
              :                  |                 :: |       :  |
Db  1664 --------KLPSPDARPTTVWNSNSTSEWVSDKSFEGRKKK-----------ENEDEEGAVNA 1707

Qy   955 AEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQG---------REGLKPLYEMQ 1007
              || :   |:     | ::  : ||  |    :||  |::           |
Db  1708 AHQGEVRAA-TERSDHLLLTPELESSNVDASSDLAT--WEGFILETTPTESEKEMANST 1763

Qy  1008 GALIKQLDAHNV----------RHNAPQPDLQSKLETL-------------DLG--EHGAEL--  1044
            ||                            | || |              |||  |
Db  1764 PVFRETIGVANVEAQPFEHSSSSSHPRVQEELTTLSGNPPSLFTDLGSGDASTGMELITAS 1823

Qy  1045 ----------LNDMKRFRDELEQSATRSVTVLGQHOGVLKSNGEINSEFKP--SPGKALVQSFN 1096
                                         |      :      |::|
Db  1824 LFTLDLESETKVKKELPSTPSPSV----------EISSSFEPTGLTPSTVL--D 1865

Qy  1097 VNRSG---QDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPN 1153
            | | |       | | ||   |:  || |   |  ||  :|  |   || :|   |
Db  1866 IEIAGVMSQTSQKTLISEISGKPTS-QSGVRDLYTGF-PMGEDFSGSFSEYPTVSVPTMK 1923

Qy  1154 DKT----ALTKSRLILDTVTIGELHELADKAKLVSDHK-PDA----------DQIKQLRQ 1198
            :  |       ::: :  |   |  |  | :|  |
Db  1924 EETVGMGGSDDERVRDTQTSSSIPTTSDNIYPVPDSKGPDSTVASSTAFPWEVMSSAEG 1983

Qy  1199 QFDTLREKRYESNEVKHYTDMGF--THNKALEANDAVKAFINAFKKEHHGVNLTTRTVL 1256
            || |||          |     |                    :          || |
Db  1984 SGEQLASVRSSVGPVLPLAVDIFSGTESPYFDEEFEEVAAVTEANERP-----TVL 2034

Qy  1257 ESQSSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPV---IPGAG 1312
            |                 :||:|| || |   :|| || |   ||    |    |
Db  2035 PTAASGNTVDLTENGYIEVNSTMSLDFPQ----------TMEPSKLWSKPEVNLDKQEIGRET 2087

Qy  1313 ITLDRAYNLSFSRTSGGLNVSFGRDGGV----------SGNIMVATGHDVMPYMTGK 1359
            |                     |:|              |: :| |  |   ||:
Db  2088 VTREKAQG----QKTFESLHSSFAPEQTILETQSLIETEFQTSDYSMLTT---LKTYITNK 2141

Qy  1360 KTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFLTEDELPGFIHGLTHGTL 1419
            |    :          :    |||            |  |:  |||     |   |
Db  2142 EVE----EEGMSIAHMSTP----------GPGIKDLESYTTHPEAPGKSHSFSATAL 2184

Qy  1420 TPAELLQKGIEHQMKQGSKLTFSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASA 1479
            |                    |                    :   |
Db  2185 V----TESGAARSVLMDSSTQEE-ESIKLFQKGVK----------------- 2214

Qy  1480 NLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFP 1539
            |  |                          :|:::|:   || |      ||
Db  2215 ------------LTNKESNADLSFSGLGSG---------------- 2236

Qy  1540 GALP 1539 (GALP)
Db  2236 1396 GALP
```

```
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW   Hydrolase; Serine protease; Cell wall; Zymogen; Signal; plasmid;
KW   Transmembrane.
FT   SIGNAL        1     33
FT   PROPEP       34    187
FT   CHAIN       188   1902       PIII-TYPE PROTEINASE.
FT   DOMAIN      188   1902       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1877   1895       MEMBRANE ANCHOR (POTENTIAL).
FT   DOMAIN     1896   1902       CYTOPLASMIC (POTENTIAL).
FT   ACT_SITE    217    217       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    281    281       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE    620    620       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DOMAIN     1867   1872       CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT                                PROTEINS.
SQ   SEQUENCE  1902 AA;  200550 MW;  87CECBAA9345F9D3 CRC64;

Query Match             2.0%;
Best Local Similarity  18.9%;    Score 193.5;  DB 1;  Length 1902;
Matches  409;  Conservative 284;  Mismatches 720;  Indels 753;  Gaps 113;

Qy    11 KAAVH--------TAAHNPVGHGVALQQG---SSSSSPQNAAASLAAEGKNRGKMPRIHQPST 62
          ||:|
Db    32 KAAISQQTKGSSLANTVTAATAKQAATDTTAATTNQAIATQLAAKGIDYNKLNKVQQQDI 91

Qy    63 AADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPG---TTHSKGATLRDLLARDDGETQ 119
          |||  |  :|:|
Db    92 YVDVIVQ----------------MSAAPASENGILRTDYSSTA-------------EIQ 121

Qy   120 HEAAAPDAARLTRSGGVKRRNMDDMAGRP--MVKGGSGEDKVPTQOKRHQLNNFGQMRQ 176
          ||:|| |::|
Db   122 QETNKVIAAQASVKAAVEQVT-QQTAGESGYGVVNGFSTKVRVVDIPK--LKQIAGVKT 177

Qy   177 TMLSKMAHPASANA------------------------GDRLQHSPPHIPGSHHEIK--EEPVGS 215
          ::|:|
Db   178 VTLAKVYYPTDAKANSMANVQAVWSNYKYKGEGTVVSVIDSGIDPTHKDMRLSDDKDVKL 237

Qy   216 TSKATTAHADRVE---------------IAQEDD---DSEFQQLHQQRLA------- 247
          | :|
Db   238 TKSDVEKFTDTVKHGRYFNSKVPYGFNYADNNDTITDDKVDEQHGMHVAGIIGANGTGDD 297

Qy   248 ----------------RERENPPQPPKLGVATPISA----------------------- 267
          |:
Db   298 PAKSVVGVAPEAQLLAMKVFSNSDTSAKTGSATVVSAIEDSAKIGADVLNMSLGSNSGNQ 357

Qy   268 -RFQPKLTAV----------------------AESVLEGTDTTQSPLKPQSML-KGSGAGVT 305
          | |:|
Db   358 TLEDPELAAVQNANESGTAAVISAGNSGTSGSATEGVNKDYYGLQDNEMVGSPGTSRGAT 417

Qy   306 PL--------AVTLDKGK-LQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQ- 352
          :|
Db   418 TVASAENTDVITQAVTITDGTGLQLGP--------ETIQLSSHDFTGSFDQK 461

Qy   353 -HLLLDNKGHLFDIKSTATSYSVLHNSHPGEIRGKLAQAGTGSVSVD-------GKS 401
          :|:
Db   462 KFYIVKDASGNL--SKGALADYT-------ADAKGKIAIVKRGEFSFDDKQXYAQAGAA 512

Qy   402 GKISL--GSGTQSHNKTMLSQPGEAHRSLLTG----IWQHPAGAARPQGESIRL--- 449
          ||:
Db   513 GLIIVNTDGTATPMTSIALTTTFPTFGLSSVTGQKLVDWVTAHPDDSL--GVKITLAML 569

Qy   450 -----HDDKI-------------HILHPELGVWQSADKDTHSQLS----- 476
          ||:
Db   570 PNQKYTEDKMSDFTSYGPVSNLSFKPDITAPGGNIWSTQNNNGYTNMSGTSMASPFIAGS 629

Qy   477 ----RQA---DGKLYA------------LKDNRTLQN 494
          :||
Db   630 QALLKQALNNKNNPFYAYYKQLKGTALTDFLKTVEMNTAQPINDINYNNVIVSPRRQGAG 689

Qy   495 LSDNKSSEKLVDKIKSYSVDQRGCVAI---LTDPGRHKMS-------IMPSLDASPE 542
          ::|:
Db   690 LVDVKAAIDALEKNPSTVVAENGYPAVELKDFTSTDKTFKLTFTNRTTHELTYQMDSNTD 749
```

```
Qy  1540 AFTSTNVS---------AALALDNRTSQSISLEL------KRAEPVTSNDISELTSTLG 1583
          :|:|                :|:  |   :|:|||:| |
Db  2237 PLPTTSVNLTDMKQIISTLYAETSHMESLGTSILGDKMEDHERMEDVSSNEVRMLISKIG 2296

Qy  1584 KHFKDSA----------TTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERY 1628
          :||                                              :|
Db  2297 SISQDSTEALDTLSHTGTEEPTTSTLPFVKLMDLERSPKQ------DPSGGKRK 2345

Qy  1629 EAVRNLKKL--VIRQQAADSHSMELGSASHST---TYNNLSRINNDGIVELLHKHFDAA 1682
          :|                |||  ||    |:
Db  2346 PKTHRPQTMSGLISNENSSASEAEEGATSPTAFLPQTYS-------VEMT-KHF--- 2391

Qy  1683 LPASSAKRLGEMMN------NDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAIL 1735
          | :                :||  || ||  :|
Db  2392 --APSESQPSDLFNVNSGEGSGEVDTLDLVYTSGTTQASSQGDSMLASHGFLEK----- 2443

Qy  1736 DGKVGREEVG---------VLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAM 1786
          :                |||::   |   |::
Db  2444 HPEVSKTBAGATDVSPTASAMFLHHSEYK-SSLVPTSTLPSTEPYKSP------SEGIED 2496

Qy  1787 SMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASALTDLKKE 1832
          :|||   || ||                :|:|:|||:
Db  2497 GLQDN---IQFE-GSTLKPSRRKTTE----------SIIIDLDKE 2527
```

```
RESULT 14
P3P_LACLC
ID   P3P_LACLC     STANDARD;      PRT;   1902 AA.
AC   P15292;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PIII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
DE   ASSOCIATED SERINE PROTEINASE).
GN   PRTP.
OS   Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG   Plasmid.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1359;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC   STRAIN=SK11;
RX   MEDLINE=89340435; PubMed=2760036;
RA   Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT   "Primary structure and organization of the gene for a procaryotic,
RT   cell envelope-located serine proteinase.";
RL   J. Biol. Chem. 264:13579-13585(1989).
CC   -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC       GROWTH OF THE BACTERIA ON MILK.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC       SUBTILASE FAMILY.
```

```
DR   EMBL; J04962; AAA03533.1; ALT_SEQ.
DR   PIR; A32634; A32634.
DR   HSSP; P00782; 2SBT.
DR   MEROPS; S08.019; -.
DR   InterPro; IPR000209; -.
DR   InterPro; IPR001899; -.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF00082; Peptidase_S8; 3.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
```

Wed Jun 6 10:19:12 2001

```
Qy  543  SHISLSLHFADAHQGLLHGKSELEAQQSVAISHGRLVVADSEGKL-FSAAIPKQGDGNELK 601
         ::  ::  : :::  ::: :    :   :    : ::
Db  750  TN-AVTSATDPNSGVLYDK-KIDGAAIKAGSNITVPAGKTAQIEFTLSLPKSFDQQQF- 806

Qy  602  MKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQ-FHPGWNLT 660
            ::                                     ::  ::
Db  807  ------------------------VEGFL------NFKGSDGSRLNLPYMGFFGDWN-- 833

Qy  661  DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGK--LHIYFDQLTKGWTGAE------ 711
          :  :   ::                    :::: :
Db  834  DGKIVDSLNGITIYSPAG------GNFGTVPLLKNKNTGTQYYGGMVTDADGNKTVDDQA 886

Qy  712  -SDCKQLKGLDGAAYLLK----------DGEVKRLNINQSTSSIKHGTENVFSLPH 757
          ::     :                   :::
Db  887  IAFSSDKNALYNDISMKYYLLRNISNVQVDILDGQGNKVTTLSSSTNRKKTYYNAHSQQY 946

Qy  758  V-RNKPEPGDALQGL---------NKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGT 806
          :                        : ::::    :
Db  947  IYYNAP----AWDGTYYDQRDGNIKTADDGSYTYRISGV-----PEGGDKR--QVFDVP 994

Qy  807  QQLERPAQTLSREGISGELKD-------IHVDHKQNLYAL------THEGEVFHQPREAW 853
          :  :: :   :           :::::                 :
Db  995  FKLDSKAPTVRHVALSAKTENGKTQYYLTAEAKDDLSGLDATKSVKTEINEVTNLDATFT 1054

Qy  854  QNGAESSSWHIKLALPQSESKLKSL--DMSHE------HKPIATFEDGSQHQLKAGGWHAY 905
          ::                        :::             :
Db  1055 DAGTTADGYTKIETPLSDEQAQALGNGDNSAELYLTDNASNATDQDASVQ--KPGSTSFD 1112

Qy  906  AAPERGPLAVGTSGSQOTVFNRLMQGVKGKVIPGSGLTVKLSAQIGGMTGAEGRKVSSKFS 965
          :                     :   :::::          :: :
Db  1113 LIVNGGGIPDKISSTTTGYEANTQG--GGTYTFSG--TYPAAVDGTYTDAQGKKHDLNTT 1168

Qy  966  ERIRAYAFNPTMSTPRPIKNAAYATQHGWQWQGREGLKPLY--EMQGALIKQLDAHNVRHNA 1023
          :                          :   ::
Db  1169 YDAATNSFTASM----PVTNADYAAQ-----VDLYADKAHTQLLKHFDT-KVRLMA 1214

Qy  1024 PQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEF 1083
          :                                        :: :
Db  1215 P----------------------TFTDLK-FNNGSDQTS------EATIKVTGTVSADT 1244

Qy  1084 KPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGE 1143
          :                                         : : ::
Db  1245 K------TVN-------------------VGHTVAALDAQHHFSVD 1265

Qy  1144 IPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQ--FD 1201
          :  :                    :         :::         ::
Db  1266 VPV---NYGDNT------IKVTATDKDGNTTTEQKTITSSYDPD----MLKKSVTFD 1309

Qy  1202 TLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQ-G 1260
          :                            ::  :
Db  1310 --------------QGVKFGTNKF------NATSAKFYDPKTG 1332

Qy  1261 SAELAKKLKN-TLLSLDS-----GESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGI 1313
          :          :  :::       ::: :::::         ::::
Db  1333 IATITGVKHPTTLQVDGKQIPIKDDLTFS------FTLDLGTLGQK-PFGVVVG--- 1381

Qy  1314 TLDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPY------MTGKKTSAGNA 1366
          :                                  :          :::::
Db  1382 --DTTQNKTFQEA------LSFIID-AVAPTLSLDSSTDAPVVTNDPNFQITG--TATDNA 1431

Qy  1367 SDWLSAKHKISPDLRI-GAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTL-TPAEL 1424
          :                                 :::
Db  1432 -QYLS----LSINGSSVASQYEDININSGK-------PG------HMAIDQPVKL 1468

Qy  1425 LQ-KGIEHQMKQGSKLIFSVDTSANLDLRAGINLNEDGSKPNGTARVSAGLSASANLAA 1483
          :                                    ::  :
Db  1469 LEGKNV-------LTVAVTDS------ED----NTTTKNITVYYEPKKTLAA 1503

Qy  1484 GSRERSTT-----------SGQFGSTTS-ASNNRPTFLNGVGAGANLTAALGVAHSSTHE 1531
          :                  :::
Db  1504 PTVTPSTTEPAQTVTLTANAAATGETVQYSADGGKTYQDVPAAGVTITA------NGTFK 1557
```

```
Qy  1532 GKPVGIF----PA--FTSTNVSAALALDNRTSQSISLEKRAEPVTSNDI--SELTSTLG 1583
          :: :                  :  :: :
Db  1558 FKSTDLYGNESPAVDYVVTNIKA---------DDPAQLQAAKQELTNLIASAKTLSASG 1607

Qy  1584 KHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDER-YEAVRNL------ 1634
          :: : ::                          ::
Db  1608 KY--DDATTTALAA-----ATQKAQTALDQTNASVDSLTGANRDLQTAINQLAAKLPAD 1659

Qy  1635 KKLVI--RQQAADSHSMELGSASHSTTYNNLSRINNDGIVEL-LHKHFDAALPASSAKR 1690
          :: :  ::
Db  1660 KKTSLLNLQSVKDALGTDLGNQTDPSTGKTFTAALDLVAQQAGTQTDQLQATLAKI 1719

Qy  1691 LGEMMNNDPALKDIIKQLSQTPFSSASVSMELKDGLREQTEKA----ILDGKVGREEVG 1745
          ::     :      :
Db  1720 LDEVL--AKLAEGIK--AATP--AEVG-NAKDAATGKTWYADIADTLTSQQASADASD 1770

Qy  1746 VL-FQDRNNLRVK---SVSVSQSVSKSEGFNTPALLGTSNSAAMSMERNIGTINFKYG 1800
          :: :                 :::
Db  1771 KLAHLQALQSLRTKVAAAVEAAKTVGKGDG-----------TTGTSDK 1808

Qy  1801 QDQNTP 1806
          :
Db  1809 GGQGTP 1814
```

RESULT 15
P2P_LACLA

```
ID   P2P_LACLA      STANDARD;      PRT;   1902 AA.
AC   P15293;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   PII-TYPE PROTEINASE PRECURSOR (EC 3.4.21.96) (LACTOCEPIN) (CELL WALL-
DE   ASSOCIATED SERINE PROTEINASE) (LP151).
GN   PRT.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OG   Plasmid pLP763.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCDO 763;
RX   MEDLINE=89311288; PubMed=2501630;
RA   Kiwaki M., Ikemura H., Shimizu-Kadota M., Hirashima A.;
RT   "Molecular characterization of a cell wall-associated proteinase gene
RT   from Streptococcus lactis NCDO763.";
RL   Mol. Microbiol. 3:359-369(1989).
CC   -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC       GROWTH OF THE BACTERIA ON MILK.
CC   -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC       SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC       E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC       PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC       ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC       INSULIN B-CHAIN.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC       SUBTILASE FAMILY.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; X14130; CAA32350.1; -.
DR   PIR; S06997; S06997.
DR   HSSP; P00782; 2SBT.
DR   MEROPS; S08.019; -.
DR   InterPro; IPR000209; -.
```

```
DR   InterPro; IPR001899; -.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   Pfam; PF00082; Peptidase_S8; 3.
DR   PRINTS; PR00723; SUBTILISIN.
DR   PROSITE; PS00136; SUBTILASE_ASP; 1.
DR   PROSITE; PS00137; SUBTILASE_HIS; 1.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW   Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW   Transmembrane.
FT   SIGNAL      1     33      POTENTIAL.
FT   PROPEP     34    187      POTENTIAL.
FT   CHAIN     188   1902      PI-TYPE PROTEINASE.
FT   DOMAIN    188   1876      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM 1877   1895      MEMBRANE ANCHOR (POTENTIAL).
FT   DOMAIN   1896   1902      CYTOPLASMIC (POTENTIAL).
FT   ACT_SITE  217    217      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE  281    281      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE  620    620      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   DOMAIN   1867   1872      CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT                             PROTEINS.
SQ   SEQUENCE 1902 AA;  200139 MW;  4B8D8B844D88CDF7 CRC64;
```

```
Query Match       2.0%;    Score 189;    DB 1;    Length 1902;
Best Local Similarity  18.8%;    Pred. No. 0.26;
Matches  394;    Conservative  251;    Mismatches 751;    Indels  698;    Gaps  100;

Qy    11  KAAVH---------TAAHNPVGHGVALQQG--SSSSSPQNAAASLAAEGKNRGKMPRIHQPST   62
             ||||          |||                 :::  ||           ||| : |
Db    32  KAAISQQTKGSSLANTVTAATAKQAATDTAATTNQAIATQLAAKGIDYNKLNKVQQQDI   91

Qy    63  AADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPCTTHSKGATLRDLLARDDGETQHEA  122
             :| ||| |||                |||||   ||  :    ::
Db    92  YVDVIVQ-----------------------MSAAPASENGTLRFDYSS-----TAEIQQET  124

Qy   123  AAPDAARLTRSGGVKRRNMDDMAGRP--MVKKGGSGEDKVPTQQKRHQLNNFGQMRQTML  179
               |||              ||             || |    :   ::  ||    :
Db   125  NKVIAAQASVKAAVEQVT-QQTAGESGYGYVNGFSTKVRVVDIPK--LKQIAGVKTVTL  180

Qy   180  SKMAHPASANA-------------------------GDRLQHSPPHIPGSIHEIK---EEPVGSTSK  218
             :| ||||                           ||    | ||:  |: | ::|||
Db   181  AKVYYPTDAKANSMANVQAVWSNYKYKGEGTVVSVIDSGIDPTHKDMRLSDDKDVKLTKS  240

Qy   219  -----ATTA-----------HADRVEIAQEDDDSEFQQLHQQRLARERENPPQPPK  258
                 |||                   :: | ||          | ||   ||||
Db   241  DVEKFTDTAKHGRYFNSKVPYGFNYADNNDTITDDTVDEQHGMHVAGIIGANGTGDDPAK  300

Qy   259  --LGVATPISARFQPKLTAVAEESVLEGTDTQSPLKPQSMLKGSGAGVTPLAVTLDKGKL  316
                || | |     ||                   |||   ::  ||   |||  |
Db   301  SVVGVA----------PEAQLLAMKVFTNSDT------SATTGSATLVSAIE--DSAKI  341

Qy   317  QLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLH  376
             ||:        :|  :|||                |||   ||:  ||  |
Db   342  GA----------DVLNMSLGSDS--------GNQTL----------EDPELAAVQ  368

Qy   377  NSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNK----------TMLSQPGEAHRSL  428
             :|:                             :|  |||
Db   369  NAN-----ESGTAAVISAGNSG--TSGSATEGVNKDYYGLQDNEMVGTPGTSRGAT  417

Qy   429  LTGIWQHPAGAARPQ--GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYAL  486
                :||            :::: |:          |   || |
Db   418  TV-----ASAENTDVITQAVTITDGTGLQLGPE--TIQLSSNDFTGSFDQK---KFVVV  466

Qy   487  KDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAIL------------TDTPGRHKMSI  533
             |||                   :::  |||:::
Db   467  KD-----ASGNLSKGKVAD---YTADAKGKIAIVKRGELTFADKQKYAQARAGAAGLII  516

Qy   534  MPS-LDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIP  592
             :: |                  |||:                           |
Db   517  VNNDGTATPVTSMALTTTFP----TFGLSSVTGQKLVD----------------WVAAHP  556

Qy   593  KQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQ  652
                                                    :::    ||
Db   593  GSKPNGTARVSAGLSASANLAAGSRERSTT-----------SGQFGSTTS-ASNNRPTF  1508
```

```
Db   557  DDSLGVKIALTLVPNQKYTE------------DKMSDF--TSYGPVSNL---SFKPDITAPGGN---  603
             :: ::|||  :| |||: |                |:      :      |      |||
Qy   653  FHPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLH---------------YFDQLT  704
                                                                            |  ||
Db   604  ------------IWSTQNNNGYTNMSGTSMASPFIAGSQALLKQALNKNNPFYAYYKQL-  651
                                                                          || |
Qy   705  KGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEP  764
                              |      | |       ||  |         :        |  |||
Db   652  ----------KGTALTDFL----KTVEMNTAQPINDINY--NNVIVSPR----  684
                                                                  ||
Qy   765  GDALQGLNKDKAQAMAVIGVNKYLALTEKG----DIRSFQIKPGTQQL  809
                                   :  ||:    :::              ||:|
Db   685  --RQGAGLVDVKAAIDALEKNPSTVVAENGYPAVELKDFTSTDKTFKLTFNRTTHELT  741
                                                           || ||| |
Qy   810  ------ERPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFHQPREAWQNGAESSS  861
                   |::|    ::  |  :     ::       ::        ||    |  :|:
Db   742  YQMDSNTDTNAVYTSATDPNSGVLYDKKIDG---AIKAGSNITVPA----GKTAQI  791
                                                                     ||
Qy   862  WHKLALPQSESKLKSLDMSHEHKPIATFE--DGSQHQLK----AGGWH---------------A  904
             ::||       ::|  :| :|  ||||   ||:|||       ||||
Db   792  EFTLSLP------KSFDQQQFVEGFLNFKGSDGSRLNLPYMGFFGDWNDGKIVDSLNGIT  845
                                                                            :
Qy   905  YA--------------------APERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVK---  944
             :||                   ||  :|||         ::  |   ::::||  :::::
Db   846  YSPAGGNFGTVPLLTNKNTGTQYYGWMVTDADGNQTVDDQAIAFSSDKNALYNDISMKY  905
             :
Qy   945  ---LSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGL 1000
                  |  | |         :    :           :|                ||| |
Db   906  LLRNISNVQVDILDGQGNKVTTLSSSTNRKKTYYNAHS----QQYYYHAPAWDG----  956
                                                       :
Qy  1001  KPLYEMQGALIKQLD--------------------AHNVRHNAPQPDLQS  1030
             |     ||| ||                       |   | ||  || |:
Db   957  -TYYDQRDGNIKTADDGSYTYRISGVPEGGDKRQVFDVPFKLDSKAPTVRHVA----LSA 1011
             :||   || |  ||          || | | |         || :|
Qy  1031  KLET------------------LDLGEHGAELLNDMKR---FRD--ELEQSATRSVTVLG  1067
             |||                   |  |||:|  |      |||                |
Db  1012  KTENGKTQYYLTAEAKDDLSGLDATKSVKTAINEVTNLDATFTDAGTTADGYTKIETPLS  1071
             :|                     || | ||
Qy  1068  QHQGVLKSNGEINSE--FKPSPGKALVQSFNVNRSQO--DL--SKSLQQAVHATPPSA  1119
                   ||             ||||                  |||
Db  1072  DEQAQALGNGDNSAELYLTDNASNATDQDASVQKPGSTSFDLIVNGGGIPDKISSTTTGY 1131
                                                          ::|
Qy  1120  ESKLQ---SMLGHFVSAGVDMSHQKGEIPLGRQRDPND-----KTALIKSRLILDTVT  1169
             |:|    | | |               | |    |       ||||| |   ||||
Db  1132  EANTQGGGTYTFSGTY-PAAVDGTYTDAQ----GKKHDLNTTYDAATNSFTASMPVTNADY 1187
             ::|                                 :|   :| | | | | | |  |
Qy  1170  IGELHELADKAKLVSDHKPDADQIKLRQQFDTLREKRYESNPVKHYTDMGFTH--NKAL 1227
                | ||       ||||  |  ||      :|:||                    ::|
Db  1188  AAQVDLYADKAH--------------TQLLKHFDT--KVRLTAPT--FTDLKFNNGSDQTS 1230
             ||                       |||              ||
Qy  1228  EANYDA---------------------VKAFINAFKKEHH-----------GVNLTTRTVLESQGSAELAK 1266
             ||||                       |||                     ||:| |   |
Db  1231  EATIKVTGTVSADTKTVNVGDTVAALDQAHHFSVDPVPVNYGDNTIKVTATDEDGNTTTEQ 1290
             ||||                                              ::
Qy  1267  K----------------------LKNTLLSLDSGESMSFSRSYGGGVSTVF----VPTLSKKVPPV--- 1307
             |                                      :  || ||         |||
Db  1291  KTITSSYDPDMLKNS--VTFDQG--VTFGANEFNATSAKFYDPKTGIATTGKVKHPTTTL 1347
             :::|                ||                    ::|| |
Qy  1308  --IPGAGITLDRAYNLSFSRTSGGLNVS-FGRDGGVSGNIMVATGHDVMPYMTGKKTSAGN 1365
                  :|||      |      |||     |                       |:
Db  1348  QVDGKQIPIKDLTFSFTLDLGTLGQKPFG-----------VVVGD---TTQNKTFQEA 1392
                                           |       ||||                |
Qy  1366  ASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELL 1425
             ||||                        :::          :|
Db  1393  LTFILDA---VAPTLSLD---SSTDAPVYTNNPFQIT------GTAT----------  1428
             ::|:                                     ||||
Qy  1426  QKGIEHOMKQGSKLTFSVDTSANLDLRAGINLNE------------------------------D  1460
             |||                                             :::
Db  1429  --DNAQYLSLSINGSSVASQYVDININSGKPGHMAIDQPVKLLEGKNVLTVAVT  1480
             ||
Qy  1461  GSKPNGTARVSAGLSASANLAAGSRERSTT-----------SGQFGSTTS-ASNNRPTF 1508
             :::          :::                              |||
```

us-09-596-784-2.rsp

Wed Jun 6 10:19:12 2001

```
Db   1481  DSEDNTTKNITVYYEPKKTLAAPTVTPSTTEPAKTVTLTANSAATGETVQYSADGGKTY 1540

Qy   1509  LNGVGAGANLTAALGVAHSSTHEGKPVGIF----PA--FTSTNVSAALALDNRTSQSISL 1562
                  :  ||   :||    ::    :     ||   ||::|
Db   1541  QDVPAAGVTVTA------NGTFKFKSTDLYGNESPAVDYVVTNIKA---------DDPA  1584

Qy   1563  ELKRAEPVTSNDI--SELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620
                :||:||:  :|  | :  :||  ||||  |||    |    :  :  :  :  |
Db   1585  QLQAAKQELTNLIASAKTLSASGKY--DDATTTALAA-----ATQKAQTALDQTNASVD  1636

Qy   1621  DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680
                :  |-  | :|:|:                           |  |:|
Db   1637  SLTG-----ANRDLQ-------------------TAINQLA------------------  1653

Qy   1681  AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEK------A 1733
                 | |||                  :  |||    |   |:     ||
Db   1654  AKLPADKKTSL------------LNQLQ------SVKAALETDLGNQTDSSTGKTFTA  1693

Qy   1734  ILDGKVGREEVGVLFQDRNNLRVKSVSVSQSVSKSEGFN--TPALLLGTSNSAA  1785
                || |   ||   :  :  :|| | : ::| :|  ||  :: ||
Db   1694  ALDDLVAQAQAGTQTDDQLQATLAKVLDAVLAKLAEGIKAATPA-EVGNAKDAA  1746
```

Search completed: June 5, 2001, 18:23:08
Job time: 364 sec

```
                  GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     June  5, 2001, 18:14:34 ; Search time 139.63 Seconds
                                      (without alignments)
                                      2117.386 Million cell updates/sec

Title:          US-09-596-784-2
Perfect score:  9448
Sequence:       1 MELKSLGTEHKAAVHTAAHN.........NPQVASALTDLKKEGLEMKS 1838

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1009251 seqs, 160854530 residues

Total number of hits satisfying chosen parameters:    1009251

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    Pending_Patents_AA_Main:*
         1:  /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
         2:  /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
         3:  /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
         4:  /cgn2_6/ptodata/2/paa/US080_COMB.pep:*
         5:  /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
         6:  /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
         7:  /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
         8:  /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
         9:  /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
        10:  /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
        11:  /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
        12:  /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
        13:  /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
        14:  /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
        15:  /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
        16:  /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
        17:  /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
        18:  /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
        19:  /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
        20:  /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
        21:  /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
        22:  /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
        23:  /cgn2_6/ptodata/2/paa/US60_COMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	9448	100.0	1838	1	PCT-US98-15426-2	Sequence 2, Appli
2	9448	100.0	1838	15	US-09-120-663-2	Sequence 2, Appli
3	9448	100.0	1838	18	US-09-412-100-28	Sequence 28, Appli
4	9448	100.0	1838	18	US-09-431-614-8	Sequence 8, Appli
5	9448	100.0	1838	19	US-09-596-784-2	Sequence 2, Appli
6	1041	11.0	201	1	PCT-US99-15425-1	Sequence 1, Appli
7	1041	11.0	201	17	US-09-350-852A-1	Sequence 1, Appli
8	268	2.8	3554	23	US-60-261-974-140	Sequence 140, App
9	257.5	2.7	10203	18	US-09-450-969-4098	Sequence 4098, Ap
10	246	2.6	3351	12	US-08-827-356-5564	Sequence 5564, Ap
11	246	2.6	3351	20	US-09-611-529-4562	Sequence 4562, Ap
12	242	2.6	2434	23	US-60-242-578-1045	Sequence 1045, Ap
13	242	2.6	2434	23	US-60-253-625-2389	Sequence 2389, Ap
14	242	2.6	2434	23	US-60-257-931-3392	Sequence 3392, Ap
15	242	2.6	2434	23	US-60-269-308-4413	Sequence 4413, App
16	240.5	2.5	2086	23	US-60-242-578-948	Sequence 948, App
17	240.5	2.5	2086	23	US-60-253-625-2292	Sequence 2292, Ap
18	240.5	2.5	2086	23	US-60-257-931-3208	Sequence 3208, Ap
19	240.5	2.5	2086	23	US-60-269-308-4231	Sequence 4231, Ap
20	238.5	2.5	2504	17	US-09-328-352-5821	Sequence 5821, Ap
21	236	2.5	3241	19	US-09-558-257-1	Sequence 1, Appli
22	234.5	2.5	1947	12	US-08-827-356-5566	Sequence 5566, Ap
23	234.5	2.5	1947	20	US-09-611-529-4400	Sequence 4400, Ap
24	233	2.5	1996	12	US-08-827-356-3129	Sequence 3129, Ap
25	233	2.5	1996	20	US-09-611-529-6335	Sequence 6335, Ap
26	232	2.5	3696	18	US-09-450-969-5942	Sequence 5942, Ap
27	232	2.5	1981	16	US-09-214-759-38	Sequence 38, Appl
28	226.5	2.4	2025	23	US-60-242-578-990	Sequence 990, App
29	226.5	2.4	2025	23	US-60-253-625-2334	Sequence 2334, Ap
30	226.5	2.4	2025	23	US-60-257-931-3282	Sequence 3282, Ap
31	226.5	2.4	2025	23	US-60-269-308-4306	Sequence 4306, Ap
32	221	2.3	3829	12	US-08-827-356-5567	Sequence 5567, Ap
33	221	2.3	3829	20	US-09-611-529-4591	Sequence 4591, Ap
34	216	2.3	2048	16	US-09-268-347-48	Sequence 48, Appl
35	213.5	2.3	2511	23	US-60-229-518-408	Sequence 408, App
36	213	2.3	1848	23	US-60-173-464-22131	Sequence 22131, A
37	211.5	2.2	1471	23	US-60-215-161-8073	Sequence 8073, Ap
38	211.5	2.2	2680	18	US-60-489-039A-7973	Sequence 7973, Ap
39	210.5	2.2	2423	23	US-60-212-411-181	Sequence 181, App
40	207.5	2.2	2142	19	US-09-540-236-3459	Sequence 3459, Ap
41	207.5	2.2	2142	23	US-60-128-476-4493	Sequence 4493, Ap
42	207.5	2.2	2478	23	US-60-242-578-1042	Sequence 1042, Ap
43	207.5	2.2	2478	23	US-60-253-625-2386	Sequence 2386, Ap
44	207.5	2.2	2478	23	US-60-257-931-3384	Sequence 3384, Ap
45	207.5	2.2	2478	23	US-60-269-308-4405	Sequence 4405, Ap

ALIGNMENTS

```
RESULT   1
PCT-US98-15426-2
; Sequence 2, Application PC/TUS9815426
; GENERAL INFORMATION:
   APPLICANT:  Cornell Research Foundation, Inc.
   TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FROM
   TITLE OF INVENTION:  ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
   NUMBER OF SEQUENCES:  5
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
     STREET:  P.O. Box 1051, Clinton Square
     CITY:  Rochester
     STATE:  New York
     COUNTRY:  U.S.A.
     ZIP:  14603
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Floppy disk
     COMPUTER:  IBM PC compatible
     OPERATING SYSTEM:  PC-DOS/MS-DOS
     SOFTWARE:  PatentIn Release #1.0, Version #1.30
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  PCT/US98/15426
     FILING DATE:
     CLASSIFICATION:
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  US 60/055,105
     FILING DATE:  06-AUG-1997
   ATTORNEY/AGENT INFORMATION:
     NAME:  Goldman, Michael L.
     REGISTRATION NUMBER:  30,727
     REFERENCE/DOCKET NUMBER:  19603/1662
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  (716) 263-1304
```

```
; TELEFAX:  (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
   SEQUENCE CHARACTERISTICS:
   LENGTH:  1838 amino acids
   TYPE:  amino acid
   STRANDEDNESS:
   TOPOLOGY:  linear
   MOLECULE TYPE:  protein
PCT-US98-15426-2
```

```
Query Match       100.08%;  Score 9448;  DB 1;  Length 1838;
Best Local Similarity   100.08%;  Pred. No. 0;
Matches 1838;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP 60

Qy   61 STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPGQOPGTTHSKGATLRDLLARDDGETQH 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPGQOPGTTHSKGATLRDLLARDDGETQH 120

Qy  121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180

Qy  181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240

Qy  241 LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300

Qy  301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360

Qy  361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVVDGKSGKISLGSGTQSHNKTMLSQ 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVVDGKSGKISLGSGTQSHNKTMLSQ 420

Qy  421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480

Qy  481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540

Qy  541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL 600

Qy  601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660

Qy  661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720

Qy  721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780

Qy  781 AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT 840

Qy  841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHLKAG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHLKAG 900

Qy  901 GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV 960

Qy  961 SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQREGLKPLYEMQGALIKQLDAHNVR 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQREGLKPLYEMQGALIKQLDAHNVR 1020

Qy 1021 HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080

Qy 1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140

Qy 1141 KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200

Qy 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260

Qy 1261 SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320

Qy 1321 LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380

Qy 1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440

Qy 1441 FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500

Qy 1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALDNRTSQSI 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALDNRTSQSI 1560

Qy 1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620

Qy 1621 DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680

Qy 1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740

Qy 1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800

Qy 1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
         ||||||||||||||||||||||||||||||||||||||
Db 1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
```

```
RESULT 2
US-09-120-663-2
; Sequence 2, Application US/09120663
; GENERAL INFORMATION:
   APPLICANT:  Bogdanove, Adam J.
   APPLICANT:  Kim, Jihyun Francis
```

```
APPLICANT: Wei, Zhong-Min
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
   ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
   STREET: P.O. Box 1051, Clinton Square
   CITY: Rochester
   STATE: New York
   COUNTRY: U.S.A.
   ZIP: 14603
COMPUTER READABLE FORM:
   MEDIUM TYPE: Floppy disk
   COMPUTER: IBM PC compatible
   OPERATING SYSTEM: PC-DOS/MS-DOS
   SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
   APPLICATION NUMBER: US/09/120,663
   FILING DATE:
   CLASSIFICATION:
PRIOR APPLICATION DATA:
   APPLICATION NUMBER: US 60/055,105
   FILING DATE: 06-AUG-1997
ATTORNEY/AGENT INFORMATION:
   NAME: Goldman, Michael L.
   REGISTRATION NUMBER: 30,727
   REFERENCE/DOCKET NUMBER: 19603/1661
TELECOMMUNICATION INFORMATION:
   TELEPHONE: (716) 263-1304
   TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
   SEQUENCE CHARACTERISTICS:
   LENGTH: 1838 amino acids
   TYPE: amino acid
   STRANDEDNESS:
   TOPOLOGY: linear
   MOLECULE TYPE: protein
US-09-120-663-2
```

```
Query Match          100.0%;  Score 9448;  DB 15;  Length 1838;
Best Local Similarity 100.0%;  Pred. No. 0;
Matches 1838;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy    1 MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP 60

Qy   61 STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQH 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQH 120

Qy  121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180

Qy  181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240

Qy  241 LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300

Qy  301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360

Qy  361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420

Qy  421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480

Qy  481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540

Qy  541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL 600
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL 600

Qy  601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660

Qy  661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720

Qy  721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780

Qy  781 AVIGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTLSREGISGELKDIHVDHKQNLYALT 840
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 AVIGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTLSREGISGELKDIHVDHKQNLYALT 840

Qy  841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG 900
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG 900

Qy  901 GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV 960
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV 960

Qy  961 SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020

Qy 1021 HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHGVLKSNGEIN 1080
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHGVLKSNGEIN 1080

Qy 1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140

Qy 1141 KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200

Qy 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260

Qy 1261 SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320

Qy 1321 LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380

Qy 1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440

Qy 1441 FSVDTSANLLDRAGINLNEDGSKPNGTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 FSVDTSANLLDRAGINLNEDGSKPNGTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500

Qy 1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALDNRTSQSI 1560
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALDNRTSQSI 1560
```

```
Db 1501 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
        ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560

Qy 1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620

Qy 1621 DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680

Qy 1681 AALPASSAKKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 AALPASSAKKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740

Qy 1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800

Qy 1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
        ||||||||||||||||||||||||||||||||||||||
Db 1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
```

RESULT 3
US-09-412-100-28
; Sequence 28, Application US/09412100
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Fan, Hao
; APPLICANT: Niggemeyer, Jennifer L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FRAGMENTS WHICH ARE
; TITLE OF INVENTION: ACTIVE BUT DO NOT ELICIT A HYPERSENSITIVE RESPONSE
; FILE REFERENCE: 21829/31 (EBC-002)
; CURRENT APPLICATION NUMBER: US/09/412,100
; CURRENT FILING DATE: 1999-10-04
; EARLIER APPLICATION NUMBER: 60/103,050
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 39
; SEQ ID NO 28
; SOFTWARE: PatentIn Ver. 2.0
; LENGTH: 1838
; TYPE: PRT
; ORGANISM: Erwinia amylovora
US-09-412-100-28

```
Query Match             100.0%;  Score 9448;  DB 18;  Length 1838;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1838;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1 MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP 60

Qy  61 STRAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQH 120
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 STRAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQH 120

Qy 121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180

Qy 181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240

Qy 241 LHQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 LHQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300

Qy 301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360

Qy 361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKIMLSQ 420
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db 361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420

Qy 421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480

Qy 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540

Qy 541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL 600
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL 600

Qy 601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660

Qy 661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720

Qy 721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780

Qy 781 AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT 840
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781 AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT 840

Qy 841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHLKAG 900
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHLKAG 900

Qy 901 GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV 960
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 901 GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV 960

Qy  961 SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020

Qy 1021 HNAPQPDLQSKLETLDIGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 HNAPQPDLQSKLETLDIGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080

Qy 1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140

Qy 1141 KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200

Qy 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260

Qy 1261 SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320

Qy 1321 LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380

Qy 1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440
```

```
Qy  1441  FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS  1500

Qy  1501  ASNRRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI  1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  ASNRRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI  1560

Qy  1561  SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK  1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK  1620

Qy  1621  DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621  DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD  1680

Qy  1681  AALPASSAAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG  1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681  AALPASSAAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG  1740

Qy  1741  REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG  1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741  REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG  1800

Qy  1801  QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS  1838
          ||||||||||||||||||||||||||||||||||||||
Db  1801  QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS  1838
```

RESULT 4
US-09-431-614-8
; Sequence 8, Application US/09431614
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; TITLE OF INVENTION: RESISTANCE
; FILE REFERENCE: 21829/41 (EBC-003)
; CURRENT APPLICATION NUMBER: US/09/431,614
; CURRENT FILING DATE: 1999-11-02
; EARLIER APPLICATION NUMBER: 60/107,243
; EARLIER FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1838
; TYPE: PRT
; ORGANISM: Erwinia amylovora
US-09-431-614-8

```
Query Match             100.0%;  Score 9448;  DB 18;  Length 1838;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1838;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy     1  MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP    60

Qy    61  STAADGISAAHQQKKSFSLRGCLGTKKFSRSRAPQGQPGTTHSKGATLRDLLARDDGETQH  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  STAADGISAAHQQKKSFSLRGCLGTKKFSRSRAPQGQPGTTHSKGATLRDLLARDDGETQH  120

Qy   121  EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS  180

Qy   181  KMAHPASANAGDRLQHSPPHIPGSSHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  KMAHPASANAGDRLQHSPPHIPGSSHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ  240

Qy   241  LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS  300

Qy   301  GAGVTPLAVTLDKGKQLAPDNPPALNTLKQTLGKDTQHYLAHHASSDGSQOHLLLDNKG  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  GAGVTPLAVTLDKGKQLAPDNPPALNTLKQTLGKDTQHYLAHHASSDGSQOHLLLDNKG  360

Qy   361  HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLSGSTQSHNKTMLSQ  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLSGSTQSHNKTMLSQ  420

Qy   421  PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD  480

Qy   481  GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVALLTDTPGRHKMSIMPSLDAS  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVALLTDTPGRHKMSIMPSLDAS  540

Qy   541  PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL  600

Qy   601  KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT  660

Qy   661  DALVIDNQLGLHHTNPEPHEILDMGHGLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  DALVIDNQLGLHHTNPEPHEILDMGHGLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG  720

Qy   721  LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM  780

Qy   781  AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT  840

Qy   841  HEGEVFHQPREAWQNGAESSSWWHKLALPQSESKLKSLDMSHEHKPIATFEDGSOHQLKAG  900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  HEGEVFHQPREAWQNGAESSSWWHKLALPQSESKLKSLDMSHEHKPIATFEDGSOHQLKAG  900

Qy   901  GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV  960

Qy   961  SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020

Qy  1021  HNAPQPDLQSKLETLDLGEHGAELLNDMKRFDRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  HNAPQPDLQSKLETLDLGEHGAELLNDMKRFDRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080

Qy  1081  SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140

Qy  1141  KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200

Qy  1201  DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTRRTVLESQG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTRRTVLESQG 1260

Qy  1261  SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320

Qy  1321  LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380
```

```
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 1838 amino acids
;           TYPE: amino acid
;           STRANDEDNESS:
;           TOPOLOGY: linear
;       MOLECULE TYPE: protein
US-09-596-784-2

Query Match          100.0%;   Score 9448;   DB 19;   Length 1838;
Best Local Similarity 100.0%;   Pred. No. 0;
Matches 1838;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     1 MELKSLGTEHKAAVHTAAHNPVGHGVALQOGSSSSSPQNAAASLAAEGKNRGKMPRIHQP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MELKSLGTEHKAAVHTAAHNPVGHGVALQOGSSSSSPQNAAASLAAEGKNRGKMPRIHQP 60

Qy    61 STAADGISAAHQOKKSFSLRGCLGTKKFSRSAPOGQPGTTHSKGATLRDLLARDDGETQH 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 STAADGISAAHQOKKSFSLRGCLGTKKFSRSAPOGQPGTTHSKGATLRDLLARDDGETQH 120

Qy   121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQOKRHQLNNFGQMRQTMLS 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQOKRHQLNNFGQMRQTMLS 180

Qy   181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGST SKATTAHADRVEIAQEDDDSEFQQ 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGST SKATTAHADRVEIAQEDDDSEFQQ 240

Qy   241 LHQORLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 LHQORLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300

Qy   301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360

Qy   361 HLFDIKSTATSYSVLNHSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 HLFDIKSTATSYSVLNHSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420

Qy   421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480

Qy   481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540

Qy   541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVADSEGKLFSAAIPKQGDGNEL 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVADSEGKLFSAAIPKQGDGNEL 600

Qy   601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660

Qy   661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720

Qy   721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780

Qy   781 AVIGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTLSREGISGELKDIHVDHKQNLYALT 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 AVIGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTLSREGISGELKDIHVDHKQNLYALT 840

Qy   841 HEGEVFHQPREAWQNGAESSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 HEGEVFHQPREAWQNGAESSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG 900
```

```
Db  1321 LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380

Qy  1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440

Qy  1441 FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500

Qy  1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560

Qy  1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620

Qy  1621 DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680

Qy  1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740

Qy  1741 REEVGVLFQDRNNLRVKSVSYSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 REEVGVLFQDRNNLRVKSVSYSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800

Qy  1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
         ||||||||||||||||||||||||||||||||||||||
Db  1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
```

```
RESULT  5
US-09-596-784-2
; Sequence 2, Application US/09596784
; GENERAL INFORMATION:
;     APPLICANT: Bogdanove, Adam J.
;     APPLICANT: Kim, Jihyun Francis
;     APPLICANT: Wei, Zhong-Min
;     APPLICANT: Beer, Steven V.
;     TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
;     TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
;     NUMBER OF SEQUENCES: 5
;     CORRESPONDENCE ADDRESS:
;         ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
;         STREET: P.O. Box 1051, Clinton Square
;         CITY: Rochester
;         STATE: New York
;         COUNTRY: U.S.A.
;         ZIP: 14603
;     COMPUTER READABLE FORM:
;         MEDIUM TYPE: Floppy disk
;         COMPUTER: IBM PC compatible
;         OPERATING SYSTEM: PC-DOS/MS-DOS
;         SOFTWARE: PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;         APPLICATION NUMBER: US/09/596,784
;         FILING DATE:
;     CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;         APPLICATION NUMBER: 09/120,663
;         FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;         NAME: Goldman, Michael L.
;         REGISTRATION NUMBER: 30,727
;         REFERENCE/DOCKET NUMBER: 19603/1661
;     TELECOMMUNICATION INFORMATION:
;         TELEPHONE: (716) 263-1304
;         TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 2:
```

```
; CURRENT APPLICATION NUMBER: PCT/US99/15425
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: 60/092,357
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Erwinia amylovora
PCT-US99-15425-1


Query Match       11.0%;   Score 1041;   DB 1;   Length 201;
Best Local Similarity  100.0%;   Pred. No. 5e-71;
Matches  201;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1  MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAAASLAAEGKNRGKMPRIHQP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAAASLAAEGKNRGKMPRIHQP 60

Qy     61  STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGPGTTHSKGATLRDLLARDDGETQH 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGPGTTHSKGATLRDLLARDDGETQH 120

Qy    121  EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180

Qy    181  KMAHPASANAGDRLQHSPPHI 201
           |||||||||||||||||||||
Db    181  KMAHPASANAGDRLQHSPPHI 201


RESULT 7
US-09-350-852A-1
; Sequence 1, Application US/09350852A
; GENERAL INFORMATION:
;  APPLICANT: Bauer, David W.
;  APPLICANT: Beer, Steven V.
;  APPLICANT: Bogdanove, Adam J.
;  APPLICANT: Collmer, Alan
;  APPLICANT: Ham, Jong Hyun
;  TITLE OF INVENTION: RECOMBINANT CONSTRUCTS AND SYSTEMS FOR SECRETION OF
;  TITLE OF INVENTION: PROTEINS VIA TYPE III SECRETION SYSTEMS
;  FILE REFERENCE: 19603/2061
;  CURRENT APPLICATION NUMBER: US/09/350,852A
;  CURRENT FILING DATE: 1999-07-09
;  PRIOR APPLICATION NUMBER: 60/092,357
;  PRIOR FILING DATE: 1998-07-10
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Erwinia amylovora
US-09-350-852A-1


Query Match       11.0%;   Score 1041;   DB 17;   Length 201;
Best Local Similarity  100.0%;   Pred. No. 5e-71;
Matches  201;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1  MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAAASLAAEGKNRGKMPRIHQP 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAAASLAAEGKNRGKMPRIHQP 60

Qy     61  STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGPGTTHSKGATLRDLLARDDGETQH 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGPGTTHSKGATLRDLLARDDGETQH 120

Qy    121  EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180
```

```
Qy    901  GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV 960

Qy    961  SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020

Qy   1021  HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080

Qy   1081  SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140

Qy   1141  KGEIPLGRQRDPNDKTALTKSRLLLDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  KGEIPLGRQRDPNDKTALTKSRLLLDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200

Qy   1201  DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260

Qy   1261  SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320

Qy   1321  LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380

Qy   1381  RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440

Qy   1441  FSVDTSANLDLRAGINLNEDGSKPNGTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  FSVDTSANLDLRAGINLNEDGSKPNGTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500

Qy   1501  ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560

Qy   1561  SLELKRAEPVTSNDISELTSTGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  SLELKRAEPVTSNDISELTSTGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620

Qy   1621  DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680

Qy   1681  AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740

Qy   1741  REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800

Qy   1801  QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
           ||||||||||||||||||||||||||||||||||||||
Db   1801  QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838


RESULT 6
PCT-US99-15425-1
; Sequence 1, Application PC/TUS9915425
; GENERAL INFORMATION:
;  APPLICANT: Cornell Research Foundation, Inc.
;  TITLE OF INVENTION: RECOMBINANT CONSTRUCTS AND SYSTEMS FOR SECRETION OF
;  TITLE OF INVENTION: PROTEINS VIA TYPE III SECRETION SYSTEMS
;  FILE REFERENCE: 19603/2062
```

```
Db  121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSEDKVPTQQKRHQLNNFGQMRQTMLS 180

Qy  181 KMAHPASANAGDRLQHSPPHI 201
        ||||||||||||||||||||
Db  181 KMAHPASANAGDRLQHSPPHI 201


RESULT   8
US-60-261-974-140
; Sequence 140, Application US/60261974
; GENERAL INFORMATION:
;  APPLICANT:  Beasley, Ellen
;  TITLE OF INVENTION:  ISOLATED HUMAN NUCLEAR HORMONE RECEPTOR
;  TITLE OF INVENTION:  PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN NUCLEAR
;  TITLE OF INVENTION:  HORMONE RECEPTOR PROTEINS, AND USES THEREOF
;  FILE REFERENCE:  CL001096-PROV
;  CURRENT APPLICATION NUMBER:  US/60/261,974
;  CURRENT FILING DATE:  2001-01-17
;  NUMBER OF SEQ ID NOS:  237
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 140
;  LENGTH: 3554
;  TYPE:  PRT
;  ORGANISM:  Human
US-60-261-974-140


Query Match          2.8%;  Score 268;  DB 23;  Length 3554;
Best Local Similarity  19.6%;  Pred. No. 1e-09;
Matches   399;  Conservative 241;  Mismatches 792;  Indels  604;  Gaps  96;

Qy   23 GHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQPSTAADGISAAHQQKK-----SF  77
Db  355 GHSEDSERHSGSAS-RNHHGSAWEQSRDGSRHPRSHDEDRASHGHSADSSRQSGTRHAET 413

Qy   78 SLRG-CLGTKKFSRSRAPQGQPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLTRSGGV 136
Db  414 SSRGQTASSHEQARSSPGERHGSGHQQSADSSRHSA-TGRGQASSAVSDRGH-RGSSGS  470

Qy  137 KRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQH 196
Db  471 QASDSEGHSENSDTQSVSGHGKAGLRQQSHQESTRGRSGE-------RSGRSGSSL-Y  520

Qy  197 SPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDSEFQOLHQQRLARE-------RE 251
Db  521 QDPEVQADKHEMRNNQ-ETAPGTQGHV---------IMKLPLRLTALDTHRWARD      565

Qy  252 NPPQPPKLGVATPISAR---------FQPKLTAVAESVLEGTDTTQSPLKPQSMLKGSG  301
Db  566 NHRGPGQVGTRDPVLARTVTVTRDTQKTLRGGLGLLPETIMDLLRSSQE-MAPD-----  617

Qy  302 AGVTPLAVT---LDKGKLQLAPDN-------------PPALNTL                329
Db  618 ---TPGPITKTELVMGTLQTAPENQALVTHRIPLVDRLHHPMNRQDVQEKDMDPATSSS  674

Qy  330 LK------QTLGKDTQHYLAHHASSDG-------SQHLLLDNKGH-----------    361
Db  675 QQTAPDTQALGTDKLHLQSETVDTEGPVVRPLTVRDIQKTQTHSQCQPMDRLGHHQQSH  734

Qy  362 -----------LFDI---KSTATSY--------SVLHNS----                378
Db  735 QESARDRSGETSGHSGSFLYQVSTHKQSESSHGWTGPSTGGRQSRHEQARDSSRHSASQ  794

Qy  379 -----HPGEIK-GKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLL    429
Db  795 DGQDTIRGHPGSSRGGRQGSHHEHSVDRSGHSGSHHSHTTSQGRSDASRGQSGSRSASRT 854

Qy  430 TGIWQHPAGAARPQGESIRLRDDKIHILHPELGVWQSADKDTHSQLSR-QADGKLYALKD  488
Db  855 TRNEEQSRDGSRHSGS--RHHEASSH----------ADISRHSQAGQGQGSEGSRTSRRQ  901

Qy  489 NRTLQNLSDNKS-SEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMP-------SLDAS  540
Db  902 GSSVSQDSDSEGHSEDSERWSGSASRNHRGSAQEQSRHGSRHPRSHEDRAGHGHSADSS   961

Qy  541 -----PESHISLSLHFADAHQ--GLLHGKSELEAQSVAISHG------RLVVADSE     583
Db  962 RQSGTPHAETSSGGQAASSHEQARSSPGERHGSRHQQSADSSRHSGIPRRQASSAVRDSG 1021

Qy  584 --GKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQ---------------ISGFF--  624
Db 1022 HWGSSGSQASDSEGHSEESDTQSVSGHGQDGPHQQSHQESARDWSGGRSGRSGSFIYQVS 1081

Qy  625 -HDD----HGQLNALVKNNFRQQH----------ACPLGNDH-QFHPGWNLTDALVIDNQ  668
Db 1082 THEQSESAHGRTRTSTGRRQGSHHEQARDSSRHSASQEGQDTIRAHPGSRR------GGR 1135

Qy  669 LGLHHTNPEPHEILDMGHLG---SLALQEGKLHYFDQLTKGWTGAESDCKQLKKGL--D   722
Db 1136 QGSHH-----EQSVDRSGHSGSHHSHTTSQGR----SDASHGQSGSRSASRQTRKDKQSGD 1187

Qy  723 GAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAV  782
Db 1188 GSRH------SGSRH--HEAASWADSSRHSQVGQEQSSGGSRTSRHQGSSV          1229

Qy  783 IGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIH--VDHKQNLYALT  840
Db 1230 ----------SQDSDSERHSDDSERLSGSASRNHHGSSREQSRDGSR             1266

Qy  841 HEGEVFHQ---------PREAWQNGAESSSWHKLALPQSESKLKSLDMSH-EHKPI      886
Db 1267 HPG--FHQEDRASHGHSADSSRQSTHHTESSS-HGQAVSSHEQARSSPGERHGSHHQQS  1323

Qy  887 ATFEDGSOHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLS  946
Db 1324 A---DSSRHSIGHGQASSAVRDSGHR-GSSGSQ-----VTNSEGHSEDSD           1365

Qy  947 AQTGGMTGAEG----------------RKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQG 996
Db 1366 TQSVSAHGQAGPHQQSHKESARGQSGESSGRSRSFLYQ--VSSHEQSESTHGQTAPSTGG 1423

Qy  997 REGLKPLYEMQGALIKOLDAHNVRHNAPQPDLQSKLETLDL--GEHGAELLNDMKRFRDE 1054
Db 1424 RQGSR------HEQARNSSRHSASQ------EGQDTIRGHPGSSRGGRQGSYH-       1464

Qy 1055 LEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHA 1114
Db 1465 -EQSVDRSGHSGYHHSHTIPQGRSDASHGQSGPRSASRQTRNEEQSG-DGSRHSGSRHHE 1522

Qy 1115 TPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLLLDTVTIGELH 1174
Db 1523 PSTRAGSSRHSQVGQGESAGSKTSRRQGS-SVSQDRDS-                      1559

Qy 1175 ELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAV 1234
Db 1560 -EGHSEDS----------ERRSESASRNHYG-                            1579

Qy 1235 KAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYG------ 1288
Db 1580 ----SAREQSRHG-SRNPRSHQEDRAS------HGHSAESSRQSCTRHAET         1619

Qy 1289 --GGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGR--DGGVSGNI 1344
Db 1620 SSGGQAA--------SSQEQARSSPGERHGSRHQQSADSSTDSG----TGRRQDSSVVGD- 1667

Qy 1345 MVATGHDVMPYMTGKKTSAGN-ASDWLSAKHKISPDLRIGAA--VSGTLQGTLQNSLK-  1399
Db 1668 -------SGNRGSSGGQASD--SEGHSEESDTQSVSAHGQAGPHQQSHQESTRGQ     1713

Qy 1400 ------------FKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTSA 1447
Db 1714 SGERSGRSGSFLYQVSTHEQSESAHGRT-GPSTGGR--QRSRHEQARDSSR--HSASQEG 1768

Qy 1448 NLDLRAGINLNEDGSKPNGVTARV-SAGLSASANLAAGSRERSTTS-GQFGSTTSASNNR 1505
Db 1769 QDTIRGHPGSSRGGRQGSHYEQSVDSSGHSGSHHSHTTSQERSDVSRGQSGSVSRQTR   1828
```

```
Qy   1506  PTFLNGVG---AGANLTAALGVAHSSTH-----EGKPVGIFPAFTSTNVSAALALDNRTSQ  1558
                :|:    :|||       |:  ::  |      .:||    : |   :::|:| |
Db   1829  NEKQSGDSRHSGSRHHEASSRADSSRHSQVGQGQSSG--PRTSRNQGSSVSQDS-DSQ     1884

Qy   1559  SISLELKRAEPVTSNDISELTSTL-----GKHFKD------SATTKMLAALKELD        1602
              :|          .:| :|      |:  |        | :  |  | :|
Db   1885  GHSEDSERWSGSASR--NHLGSAWEQSRDGSRHPGSHHEDRAGHGHSADSSRQSGTRHTE   1942

Qy   1603  DAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADS--HS-MELGSASHSTT   1659
              :  |    ::        :|        ||     ::  :||      .| ||  :
Db   1943  SSSRCQ--AASSHEQARSSAG-ERHGSHHQL------QSADSSRHSGIGHGQAS----    1987

Qy   1660  YNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVS   1719
                :|         ||          |||              :|       :
Db   1988  ----SAVRDSG---HRGYSGSQASDSE--GHSEDSD--------TQSVS             2019

Qy   1720  MELKDGLREQTEKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPA      1775
              :|               :|       :| :|        :|: |:    :|:
Db   2020  AQGKAGPHQQSHKESARGQSGES-----SGRSGSFLYQVSTHEQSESTHQSAPS        2069


RESULT 9
US-09-450-969-4098
;  Sequence 4098, Application US/09450969
;  GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: PATH99-09A
;  CURRENT APPLICATION NUMBER: US/09/450,969
;  CURRENT FILING DATE: 1999-11-29
;  NUMBER OF SEQ ID NOS: 7544
;  SEQ ID NO 4098
;  LENGTH: 10203
;  TYPE: PRT
;  ORGANISM: S. epidermidis
US-09-450-969-4098

Query Match             2.7%;  Score  257.5;  DB 18;  Length 10203;
Best Local Similarity  17.7%;  Pred. No. 4.7e-08;
Matches  396;  Conservative  319;  Mismatches 805;  Indels 723;  Gaps  90;

Qy     33  SSSSPQNAAASL-AAEGKNRGKMPRIHQPSTRAADGISAAHQQKKSFSLRGCLGTKKFSRS    91
             :: |||    |      |     |:  ::| |  ||  .|:|    :::| :|
Db   6262  ATQSVQNAEQALHGAEKLNQDK-----QTSSTELDGLTDLTDAQRE-KLREQINTSNSRDD  6316

Qy     92  APQ--GQPGTTHSKGATLRDLLARDDG-----ETQHEAAAPDAAR----------       129
            |    |.|  ::|     : ::|            :  |   || :
Db   6317  IKQKIEQAKALNDAMKKLKEQVAQKDGVHANSDYTNEDSAQKDAYNNALKQAEDIINNS    6376

Qy    130  -----LTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSK      181
                   :|          |:      :|:       :|:  :     :|
Db   6377  NPNLNAQDITNALNNIKQAQDNLHGAQKLQ-----QDKNTTNQAIGNLNHLNQPQKDALIQ  6432

Qy    182  MAHPASA----------NAGDRLQHSPPHIPGSHHEIK---------               209
             |  |:             |  :|   |     .|
Db   6433  AINGATSRDQVAEKLKEAEALDEAMKQLEDQVNQDDQISNSSPFINEDSDKQKTYNDKIQ   6492

Qy    210  --EEPVGSTSKAT------------TAHADRVEIAQEDDSEFQQLHQQRLARE          249
               |  |.:| |             ||  :: |  |::|:  :|| | :|
Db   6493  AAKEIINQTSNPTLDKQKIADTLQNIKDAVNNLHGDQ-KLAQSKQDANNQLNHLDDLTEE   6551

Qy    250  RENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAV   309
                |       |:  |         |                             |:
Db   6552  QKN-----------HFKP------LINNADTRDEVNK----------               6571

Qy    310  TLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTA   369
            :|:  |:|        | |  .: :|:    |    | :|   |:|:|||  :|||: |
Db   6572  -----QLEIAKQLNGDMSTLHKVINDKQIQHLSNYINADNDKKQNYDN------AIK      6618

Qy    370  TSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLL   429
```

```
Db   6619  EAEDLIHN-HPDTLDHKALQ----------DLLNKIDQAHNEL-                    6650
            :|: .||  |.|||||||                :||:|||| |
Qy    430  TGIWQHPAGAAR----------POGESIR-LHDDKIHILPELGVWQS                 466

Db   6651  ------NGESRFKQALDNALNDIDSLNSLNVPQRQTVKDNINHVTLESLAQEL          6698
                  || :|  |:| |  |:  :| |         ::|: :|| |:: | |:
Qy    467  ADKDTHSQLSRQADGKLYALKD----------NRTLQNLSDNKSSEKLVDKI-           508

Db   6699  ----QKAKELNDAMKAMRDSIMNQEIRKNSNYTNEDLAQQNAYNHAVDKINHIIGE        6751
                      | :       :: |:.: |||||   || .||  |    |||  |  |||
Qy    509  ----------KSYSVDQRGQVAILTDTPGRHKMSIMPSLD                        538

Db   6752  DNATMDPQIIKQATQDINTAINGLNGDQKLQDAKTDAKQQIINFTGLTEPQKQAL-----   6806
                .|        |  :|       :|      ::::| |  :   :|:|:|  ||
Qy    539  ASPESHISLSLHFADAHQGLLHGK--SELEAQSVAISHGRLVVAD---SEGKLF         587

Db   6807  --ENIINQQTSRANVAKQLSHAKFLNGKMEELKVAVAKASLVRQNSNYINEDVSEKEAY   6863
              |:       ||   ||  : | |    |: | |:     |:| || |||   |:||
Qy    588  SAAIPKQGDGNE-LKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACP   646

Db   6864  EQAIAK---GQEIINSENNPTISSTDINRTIQEINDAEQNLHGD----NKLRQAQEI     6913
            ||| |    |:||:|:|||||:||||:|||:|:||||:||||||    |||| |||:
Qy    647  LGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKG   706

Db   6914  AKNEIQNLDGLNSAQITKLIQDIGRTTTKPAVTQKLEEAKAINQAMQQLKQSIADK----  6969
             ||||:|| |||:|| |:|||||||| |||||||||||||| |||||||| |: ||:
Qy    707  WTGAESDCKQLKKGLDGAAYLLKDGEVKRLNIN-----QSTSSIKHGTENVFSLPHVRN    760

Db   6970  ----DATLNSSNYLNEDSEKKLAYDNAVSQAEQLINQLNDPTMDISNIQAITQ          7018
                || |||||| ||||||||||  ||||:||||||||||| || :::: | |
Qy    761  KP-EPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGD----IRSFQIKPGTQOLERPA     813

Db   7019  KVIQAKDSLHGANKLAQNQADSNLIINQSTNLNDKQKQALNDLIINHAQTKQQVAEIIAQA  7078
            : :||:||||||||| ||||:|||||| |||||||:| |||||||||||| |||||||| |
Qy    814  QTLSREGISGELKDIHVD----HKQNLYALTHEGEVHQPREAWQNGAESSSW-----      862

Db   7079  NKLNNE--MGTLKTLVEEQSNVHQSKY-INEDPQVQNIYNDSIQKGREILNGTTDDVLN   7135
            |||:|   |||||||||||||||||||  ||| ||||| ||| |||||||||||||||:
Qy    863  -HKLALPQSESKLKSLDMSHEHKPIATFEDGSQH----------QLKAGGWHAYAAPERG   911

Db   7136  NNKIADAIQNIHLTKNDLHGDQKLQKAQQDATNELNYLTNLNNSQRQSEHDEINSAPSRT  7195
            ||| |||:|||  ||||||||||||:||||||||||||||||||| :|| ||||| ||:
Qy    912  PLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAY   971

Db   7196  EVSNDLNHAKAL-NEAMRQLENEVALENSVK-KLSDFIN----EDEAAQNEYSNALQKA   7248
            ||:| |||| |  ||||||||||| ||||:  |||| |     |||| || ||| | |
Qy    972  --AFNPTMST--PRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLD---AHN----   1018

Db   7249  KDIINGVPSSTLDKATIEDALLELQNARESLHGEQKLQEAKNQAVAEIDNLQALNPGQVL  7308
            |||||| || ||||| |||||||||||||||||||||||||| ||||||||||||||||
Qy   1019  -----VRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFERDELEQSATRSVTVLGQHQGVL 1073

Db   7309  AEKTLVNQASTKPEVQEAL-----QKAKELNEAMKALKTEINKK--------EQIK     7351
            |||||||||||||||||||     |||||||||||||| ||||         ||||
Qy   1074  KSNGEINSEFKPSPGKALVQSFN-VNRSGQ-----DLSKSLQQAVHATPPSAESKL     1123

Db   7352  ADSRYVNAD--------SGLQANYNSALNYGSQIIATTQPPELNKDVINRATQTIKTAENNL 7405
             |:           | ||| |||| || |||||||||||| ||||:||| |:|| |||||
Qy   1124  --QSMLGHFVSAG-VDMSHQKGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKA   1180

Db   7406  NGQSKLAEAKSDGNQSIEHLQG----------LTQSOK----DKQHDLLNQA         7443
            |||| | | ||||||| |||||          |||| |    ||| ||||||
Qy   1181  KLVSDHKPDADQI----KQLRQOFDTLREKRYESNPVKHYTD----MGFTHNKALEA    1229

Db   7444  QT----KQQVDDIVNNSKQLDNSMNQLQQIVNNDNTVKQNSDFINEDSSQQDAYNHAIQA  7499
            |     ||| | | || |||||| ||||| |||| | |||| |||| | ||| ||| | |
Qy   1230  NYDAVKAFINAFKKEH----------HGVNLTTR----------                   1253

Db   7500  AKDLITAHPTIMDKNQIDQAIENIKQALNDLHGSNKLSEDKKEASEQLNLSLTNGQKD     7559
            ||||:|||| ::|||||||||||||||||||||||||| |||| | ||||| |||||
Qy   1254  TVLESOGSAEL----------AKKLKNTLLSL-----DSGESMSFSRSYGGVSTVFVPT   1298
```

```
  SOFTWARE:  PatentIn
CURRENT APPLICATION DATA:
  APPLICATION NUMBER:  US/08/827,356
  FILING DATE:  01-APR-1997
PRIOR APPLICATION DATA:
  APPLICATION NUMBER:  60/014,477
  FILING DATE:  01-APR-1996
  APPLICATION NUMBER:  60/016,743
  FILING DATE:  02-MAY-1996
  APPLICATION NUMBER:  60/020,016
  FILING DATE:  14-JUN-1996
INFORMATION FOR SEQ ID NO:  5564:
  SEQUENCE CHARACTERISTICS:
    LENGTH:  3351 amino acids
    TYPE:  amino acid
    TOPOLOGY:  linear
  MOLECULE TYPE:  protein
  HYPOTHETICAL:  YES
  ORIGINAL SOURCE:
    ORGANISM:  Staphylococcus aureus
  FEATURE:
    NAME/KEY:  misc_feature
    LOCATION:  1...3351
US-08-827-356-5564
```

```
Query Match                 2.6%;  Score 246;  DB 12;  Length 3351;
Best Local Similarity       18.0%;  Pred. No. 4.5e-08;
Matches 390;  Conservative 299;  Mismatches 797;  Indels 676;  Gaps  94;

Qy    11 KAAVHTAAHNPVGHGVALQQGSSSSSPQ--NAAASLAAEGKNRGKMPRIHQPSTAADG-  66
            : ::  :  :  :  :::     :::::    :   :       :    ::
Db   214 KQGAYTDAYNAAKNIV----------------NGSPNVITNAADVTAA--------TQRVNNAETSLNGD 259

Qy    67 --ISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQ---HE 121
           ||  |  ::              |::      |      :|  |
Db   260 TNLATAKQQAKD-ALR--------------------QMTHLSDAQKQSITGQIDSATQVTGVQ 301

Qy   122 AAAPDAARLTRSGGVKRR--NMDDM-AGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQT 177
            :  |  :: ||   | |    ||| |  |     |  |: |
Db   302 SVKDNATNLDNAMNQLRNSIANKDEVKASQPYVDADT----DKQNAYNTAVTSAEN 353

Qy   178 MLSKMAHP-------ASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADR----VEI 229
          :: :::|     :        :: : : :  : :    :::|  ||::    ::|
Db   354 IINATSQPTLDPSAVTQAANQVNTNKTALNGAQN-------LANKKQETTANINRLSHLNN 407

Qy   230 AQEDD---------DSEFQQLHQ--QRL--ARERENPPQPPKLGVATPISA 267
          |::|        ::   |  |   |:    |
Db   408 AQKQDLNTQVTNAPNISTVNQVKTKAEQLDQAMERLINGIQDKDQVKQSVNFTDADPEKQ 467

Qy   268 RFQPKLTAVAESVL--EGTDTTQSPLKPQSMLKGSGAGVTPLAVTLDKGKLQLQAPDNPP 324
          ||::| ::   | ||    |:  ||
Db   468 TAYNNAVTAAENIINQANGTNANQSQVE---------------------------A 496

Qy   325 ALNTL--LKQTLGKDTQHYLAHHASSDGSQHL-LLDNKGHLFDIKSTATSYSVLHNSHPG 381
           |::|   ::::    ::|   :  :|  :|| ||| :: |
Db   497 ALSTVTTTKQALNGDRK---VTDAKNNANQTLSTLDN----------------LNNAQKG 537

Qy   382 EIKGKLAQAGT-----------------GSVSVDGKSGKISLGSGT------QSHN 414
          ||| :|  |:                  || | :: :: ||
Db   538 AVTGNINQAHTVAEVTQAIQTAQELNTAMGNLKNSLNDKDTTLGSQNFADADPEKKNAYN 597

Qy   415 KTMLSQPGEAHRSLLTGIWQHPAGAARPQG------------ESIRLH----DDKIHI 456
          |::|| |      ::     |   :  |:|            ::|:|    |:::|
Db   598 EAVRNAENILNKSTGNVPKDQVEAAMNQVNTTKAALNGTQNLEKAKQHANTAIDGLSHL 657

Qy   457 LHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQR 516
          |||  | :: :|   | :  :| ||  |:|| ||  |:  :::  |
Db   658 ------------TNAQKEALKQLVQQSTTVAEAQGNEQKANNV--DAAMDKLRQSIADNATTKQ 707

Qy   517 GQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAH------QGLLHGKSELEAQS 569
           | :::||   :       ||   ||   |: ||||          :::  ::   ::
Db   708 NQ---------NYTDASPNK--------KDAYNNAVTTAQGIIDQTTNPSLDP 743
```

```
Db  7560 TILNHIFSAPTRSQVGEKIASAKQLNNTMKALRDSIADNNEILQSSKYFNEDSEQ---- 7614
         :  |  :  :|:|   ||: :||::|  | :   |::     |::    | |
Qy  1299 LSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTG 1358
                      ||| :  ::
Db  7615 ----------------QNAYNQAVNKAKNIIN--------DQPTPVMANDEIQSVLNE 7648

Qy  1359 KKTSAGN------ASDWLSAKHKISPDLRIGAAVSGTLQGTIQNSLK------------ 1399
         | |:||   : ||  |:|  |  :    ||  | |: | :||| :||
Db  7649 VKQTKDNLHGDQKLANDKTDAQATLNALNYLNQAQRGNLETKVQNSNSRPEVQKVVQLAN 7708

Qy  1400 -----FKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTS--ANLD--- 1450
            | | :   |  |:  |   ||  :|    |::: : |: |  ::
Db  7709 QLNDAMKKLDDALTG------------NDAIKQTSNY-INEDTSQQVNFDEY 7747

Qy  1451 LRAGINLNEDGSKPNGVTARVSA--------GLSASANLAAGSRERSTTSGQFGSTTS 1500
          | |:|:|::     ::          :
Db  7748 TDRGKNIVAEQTNPNMSPTNINTIADKITEAKNDLHGVQKLKQAQQQSINFINQMTGLNQ 7807

Qy  1501 ASN----------------NRPTFLNGV--GAGANLTAALGVAHSSTHEGKPVGIF 1538
            :             |: ||     |:| :   | :|:   | :: ||
Db  7808 AQKEQLNQEIQQTQTRSEVHQVINKAQALNDSMNTLRQSITDEHEVKQTSNYINETVGNQ 7867

Qy  1539 PAFTSTNVSAALALDNRTSQSI--SLELKRAEPVTSN-DISELTSTLGKHFKDSATTKML 1595
         | | : |  | |:|    |:|   :::   |:  | |    : : | |::|     |
Db  7868 TAYNNA-VDRVKQIINQTSNPTMNPLEVERA--TSNVKISKDALHGERELNDNKNSKTF 7923

Qy  1596 AA--LKELDDAKPAEQLHILQQHF---------SAKDVVG---------- 1624
           :  | : :     | ::          :| ||
Db  7924 AVNHLDNLNQAQKEALTHEIEQATIVSQVNNIYNKAKALNNDMKKLKDIVAQQDNVRQSN 7983

Qy  1625 -------------------DERYEAVRNLKKLVIRQQAAD-SHSM 1649
                             | :|| |::| | :::|  :|||  |:|
Db  7984 NYINEDSTPQNMYNDTINHAQSIIDQVANPTMSHDEIENAINNIKHAI---NALDGEHKL 8040

Qy  1650 ELGSASHSTTYYNLSRIN--NDGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIK 1706
         :|  |::   :||:|:|   |  :||::: ||   |: :||::::   ::| | |::|
Db  8041 QQAKENANLLINSLNDLNAPQRDAINRLVNE--------AQTREKVAEQLQSAQALNDAMK 8093

Qy  1707 QLQSTPFSSASVSMELK-----DGLREQTEKAILDGKVGREEVGVLFQ---DRNNLRV 1756
         ||:: || ::| :: |       |:||  :| | |:|  |:   || |   | :||:|
Db  8094 HLRNSIQNQSSVRQESKYINASDAKKEQYNHAV-----REVENIINEQHPTLDKEIIKQ 8147

Qy  1757 KSVSVSQSVSKSEGFNTPALLLGTSNSAAMS-------MERNIGTINFKYGQDQNTPRRFTLE 1812
         |||  | | ::    | :|   :||::| :        |   |  |::| |::||:  :|
Db  8148 LTDGVNQA---NNDLNGVELLDADKQNAHQSIPTLMHLNQAQQNALNEKINNAVTRTEVA 8204

Qy  1813 GGIAQANPQVASALTDLKKEGLE 1835
         | :| |    |||  | ||:
Db  8205 AIIGQAK------LLDHAMENLE 8221
```

```
RESULT 10
US-08-827-356-5564
; Sequence 5564, Application US/08827356
; GENERAL INFORMATION:
;   APPLICANT:  George H. Shiner, Jr.
;   APPLICANT:  George H. Miller
;   APPLICANT:  Roberta S. Hare
;   APPLICANT:  Karen J. Shaw
;   TITLE OF INVENTION:  STAPHYLOCOCCUS AUREUS RELATED
;   TITLE OF INVENTION:  COMPOSITIONS AND METHODS
;   NUMBER OF SEQUENCES:  5574
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Schering-Plough Corporation
;     STREET:  2000 Galloping Hill Road
;     CITY:  Kenilworth
;     STATE:  New Jersey
;     COUNTRY:  USA
;     ZIP:  07033-0530
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
Qy  1438 KLTFSVDTSANLD----------------------------------LRAGINLNEDGSKPNGVTARVSAGLSASANL 1481
                                                          ::  :  :| |  : :|   |:
Db  1692 ------TATSLDQAMDQLSQAINDKDQILADGNYLNADPDKQNAYKQAV-AKAEALLNK 1743

Qy  1482 AAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAA----------LGVAHSSTHEGK 1533
              :                                          |||
Db  1744 QSGTNE--VQAQVESITNEVNAAKQALNG--NDNLANAKQQAKQQLANLTHLNDAQKQ 1797

Qy  1534 PVGIFPAFTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSAT- 1591
                                                               :|  :||
Db  1798 -----SFESQITQAPLVTDVTTINQKAQTLDHAMELLRNSVADNQTTLASEDYHDATAQ 1851

Qy  1592 ------TKMLAALKEL------------DDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLK 1635
                   :||                  ||
Db  1852 RQNDYNKAVTAANNIINQTTSPTMNPDDVNGA--TTQVNNTKVALLGDENLAAAK-- 1904

Qy  1636 KLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMM 1695
              | :::                                    ||
Db  1905 -----QQANNRLDQLDHLNNAQKQQLQSQITQSSDIAAVNGHKQTAESLNTA--MGNLi 1956

Qy  1696 NNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVGRE----------EV 1744
          :::                                                        |:
Db  1957 N----AIADHQAVEORGNFINADTDKQTAYNTAVNEAAAMINKQTGQNANQTEVEQAITKV 2013

Qy  1745 GVLFQDRN----NLRVKSVSVSQSVSKSEGFNTP----------ALLLGTSNSAAMSMERNIGT 1794
               :                                     :|  :|  :||
Db  2014 QTTLQALNGDHNLQVAKTNATQAIDALTSLNDPQKTALKDQVTAATLVTAVHQIEQNANT 2073

Qy  1795 IN 1796
            :|
Db  2074 LN 2075
```

```
Qy  570  VAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHG 629
         |:|:|  | ||
Db  744  TVINQAGQVSTSKNAL-------NGNE-NLEAAKQQAT-----------------QSLG 778

Qy  630  QLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILLDMGHLGS 689
          | |:: | : ::
Db  779  SLDTL--NNAQKQ-----------------AVTNQINGAHTVDEANQIKQNAQNLN 815

Qy  690  LALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQST------ 742
                |||  :: :
Db  816  TAMGNLKQAIADKDATKATVNFTDADQAKQQAYNTAVTNAENIISKANGGNATQTEVEQA 875

Qy  743  ------SSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKY 788
               :::::                               ||  |||
Db  876  IQQVNAAKQALNGNANVQHAKDEATAL--INNSNDLNQAQKDALKQQVQNATTVAGVN-- 931

Qy  789  LALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKD------IHVD-HKQNLY----- 837
          :::                                         ||| |||
Db  932  -----NVKQTAQELNN-AMTQLKQGIADKEQTKADGNFVNADSDKQNAYNQAV 978

Qy  838  -ALTHEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLLDMSHEHKPIATFEDG 892
          ::  ||
Db  979  AKAEALISGTPDVVVTPSEI------TAALNKVT-QAKNDLNG------NTNLATAKQN 1024

Qy  893  SQH-------QLKAGGW----HAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKV-IP 937
          :|:                                               |||  ||
Db  1025 VQHAIDQLPNLNQAQRDEYSKQITQATLVPNVNAIQQAATTLNDAMTQLKQGIANKAQIK 1084

Qy  938  GS-----GLIVKLSAQTGGWTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQH 992
          ||                                                   ||
Db  1085 GSENYHDADTDKQTAYDNAVTKA---------EELLKQTTNPTMD-PNTIQQALTKVND 1133

Qy  993  GWQGREGLKPLYE----------MQGALIKQLD--------AHNVRHNAPQPD 1027
                                  ||  ||:          :||
Db  1134 TNQALNGNQKLADAKDAKTLGTLDHLNDAQKQALTTQVEQAPDIATVNNVKQNAQNLN 1193

Qy  1028 -----LQSKLETL----------DLGEHGAELL---NDMKRFRDELEQS 1058
          ||  |:||                 : ::: |||
Db  1194 NAMTNLNNALQDKTETLNSINFTDADQAKKDDYTNAVSHAEGILSKANGSNASQTEVEQA 1253

Qy  1059 ATRSVTVLGQHOGVLKSNGEIN---------SEFKPSPGKALVQSF-------NVN 1098
                                        |                      :|:
Db  1254 MQR-----VNEAKQALNGNDNVQRAKDAAKQVITNANDLNQAQKDALKQQVDAAQTVANVN 1309

Qy  1099 ---RSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVD---------MSHQKGEIPLG 1147
          :: |||::                                   |||: :|:   |||
Db  1310 TIKQTAQDLNQAMTQLKQGIADKDQTKAN---GNFVNADTDKQNAYNNAVHAEQII--- 1363

Qy  1148 RQRDPNDKTALTKSRLLLDTV-TIGEL--HEL--ADKAKLVSDHKPDADQ-------- 1192
                |                :|            : :|
Db  1364 -SGTPNANVDPQQVAQALQQVNQAKGDLNGNHNLQVAKDNANTAIDQLPNLNQPQKTALK 1422

Qy  1193 ----------IKQ---LRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAV 1234
                   ||:                               ||
Db  1423 DQVSHAELVTGVNAIKQNADALNNAMGTLKQQIQANSQVPQSVD--FT----QADQDKQ 1475

Qy  1235 KAFINAFKKEHHGVNLTRTVLES----------QGSAELAKKLKNTLLSLDS 1277
          :| ::||::  ::  :| : ::|:
Db  1476 QAYNNAANQAQQIANGTPTPVLAPDTVTKAVITMNQKDALNGDEKLAQAKQDALANLDT 1535

Qy  1278 GESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLS----FSRTSGGLNVS 1333
              |                                         |||
Db  1536 LRDLN------QPQRDALRNQINQAQALATVEQTKQNAQNVNTA 1573

Qy  1334 FGR-DGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWL---SAKHKISPDLRIGAAV 1386
           :|                                  |||
Db  1574 MGNLKQGIANKDTVKASENYHDADVDKQTAYTNAVSQAEGIINQTTNPLNPDDITRALT 1633

Qy  1387 SGT-LQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIE-------HQMKQGS 1437
           :|                                  :|||      |||:
Db  1634 QVTDAKNSLNGEAKLATEKQNAKDAVSGMTHLNDAQKQALKGQIDQSPEIATVNQVKQ-- 1691
```

RESULT 11
US-09-611-529-4562
; Sequence 4562, Application US/09611529
; GENERAL INFORMATION:
; APPLICANT: George H. Shimer,Jr.
; APPLICANT: George H. Miller
; APPLICANT: Roberta S. Hare
; APPLICANT: Karen J. Shaw
; TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods
; FILE REFERENCE: 1034/1C963US1
; CURRENT APPLICATION NUMBER: US/09/611,529
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/417,811
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/353,718
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 09/266,557
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,556
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,555
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,542
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/266,541
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/037,934
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 09/036,720
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: US 09/036,338
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,334
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,221
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,137
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: US 09/036,082
; PRIOR FILING DATE: 1998-03-06

```
;  PRIOR APPLICATION NUMBER:  US 09/036,081
;  PRIOR FILING DATE:  1998-03-06
;  PRIOR APPLICATION NUMBER:  US 09/036,079
;  PRIOR FILING DATE:  1998-03-06
;  PRIOR APPLICATION NUMBER:  US 09/035,913
;  PRIOR FILING DATE:  1998-03-06
;  PRIOR APPLICATION NUMBER:  US 09/035,744
;  PRIOR FILING DATE:  1998-03-06
;  PRIOR APPLICATION NUMBER:  US 08/827,356
;  PRIOR FILING DATE:  1997-04-01
;  PRIOR APPLICATION NUMBER:  US 08/831,156
;  PRIOR FILING DATE:  1997-04-01
;  PRIOR APPLICATION NUMBER:  US 60/014,477
;  PRIOR FILING DATE:  1996-04-01
;  PRIOR APPLICATION NUMBER:  US 60/016,743
;  PRIOR FILING DATE:  1996-05-02
;  PRIOR APPLICATION NUMBER:  US 60/020,016
;  PRIOR FILING DATE:  1996-06-14
;  NUMBER OF SEQ ID NOS: 7451
;  SEQ ID NO 4562
;  LENGTH: 3351
;  TYPE: PRT
;  ORGANISM: Staphylococcus aureus
US-09-611-529-4562
```

```
Query Match      2.6%;   Score 246;   DB 20;   Length 3351;
Best Local Similarity   18.0%;   Pred. No. 4.5e-08;
Matches   390;   Conservative 299;   Mismatches 797;   Indels 676;   Gaps 94;

Qy    11 KAAVHTAAHNPVGHGVALQQGSSSSSPQ--NAAASLAAEGKNRGKMPRIHQPSTAADG- 66
Db   214 KQGAYTDAYNAAKNIV------NGSPNVITNAADVTAA--------TQRVNNAETSLNGD 259

Qy    67 --ISAAHQQKKSFSLRGCLGTKKFSRSRSAPQGOPGTTHSKGATLRDLLARDDGETQ--HE 121
Db   260 TNLATAKQQAKD-ALR---------------QMTHLSDAQKQSITGQIDSATQVTGVQ 301

Qy   122 AAAPDAARLTRSGGVKRR---NMDDM-AGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQT 177
Db   302 SVKDNATNLDNAMNQLRNSIANKDEVKASQPYVDADT------DKQNAYNTAVTSAEN 353

Qy   178 MLSKMAHP--------ASANAGDRLQHSPPIHPGSHHEIKEPVGSTSKATTAHADR---VEI 229
Db   354 IINATSQPTLDPSAVTQAANQVNTNKTALNGAQN---------LANKKQETTANINRLSHLNN 407

Qy   230 AQEDD------------DSEFQQLHQ--QRL--ARERENPPQPPKLGVATPISA 267
Db   408 AQKQDLNTQVTNAPNISTVNQVKTKAEQLDQAMERLINGIQDKDQVKQSVNFTDADPEKQ 467

Qy   268 RFQPKLTAVAESVL---EGTDTTQSPLKPQSMLKGSGAGVTPLAVTLDKGKLQLAPDNPP 324
Db   468 TAYNNAVTAAENIINQANGTNANQSQVE-------------------A 496

Qy   325 ALNTL--LKQTLGKDTQHYLAHHASSDGSQHL-LLDNKGHLFDIKSTATSYSVLHNSHPG 381
Db   497 ALSTVTTTKQALNGDRK---VTDAKNNANQTLSTLDN-------------LNNAQKG 537

Qy   382 EIKGKLAQAGT---------------GSVSVDGKSGKISLGSGT------QSHN 414
Db   538 AVTGNINQAHTVAEVTQAIQTAQELNTAMGNLKNSLNDKDTTLGSQNFADADPEKKNAYN 597

Qy   415 KTMLSQPGEAHRSLLTGIWQHPAGAARPQG---------ESIRLH---DDKIHI 456
Db   598 EAVRNAENILNKSTGTNVPKDQVEEAAMNQVNTTKAALNGTQNLEKAKQHANTAIDGLSHL 657

Qy   457 LHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQR 516
Db   658 --TNAQKEALKQLVQQSTTVAEAQGNEQKANNV--DAAMDKLRQSIADNATTKQ 707

Qy   517 GQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAH--------QGLLHGKSELEAQS 569
Db   708 NQ-----------------NYTDASPNK----------KDAYNNAVTTAQGIIDQTTNPSLDP 743
```

```
Qy   570 VAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHG 629
Db   744 TVINQAAGQVSTSKNAL-------------------NGNE-NLEAAKQQAT----------QSLG 778

Qy   630 QLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHLGS 689
Db   779 SLDTL--NNAQKQ-------------------AVTNQINGAHTVDEANQIKQNAQNLN 815

Qy   690 LALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQST-------- 742
Db   816 TAMGNLKQAIADKDATKATVNFTDADQAKQQAYNTAVTNAENIISKANGGNATQTEVEQA 875

Qy   743 ---------------SSIKHGTENVFSLPHVRNKPEPGDALQGLNKDKAQAMAVIGVNKY 788
Db   876 IQQVNAAKQALNGNANVQHAKDEATAL--INNSNDLNQAQKDALKQQVQNATTVAGVN-- 931

Qy   789 LALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKD-------IHVD-HKQNLY 837
Db   932 --NVKQTAQELNN-AMTQLKQGIADKEQTKADGNFVNADSDKQNAYNQAV 978

Qy   838 --ALTHEGEVFHQPREAWQNGAESSWHKLALPQSESKLKSLDMSHEHKPIATFEDG 892
Db   979 AKAEALISGTPDVVVTPSEI-------TAALNKVT-QAKNDLNG------NTNLATAKQN 1024

Qy   893 SQH---------QLKAGGW------HAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKV-IP 937
Db  1025 VQHAIDQLPNLNQAQRDEYSKQITQATLVPVNAIQQAATTLNDAMTQLKQGIANKAQIK 1084

Qy   938 GS--GLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQH 992
Db  1085 GSENYHDADTDKQTAYDNAVTKA----EELLKQTTNPTMD-PNTIQQALTKVND 1133

Qy   993 GWQQREGLKPLYE------MQGALIKQLD----AHNVRHNAPQPD 1027
Db  1134 TNQALNGNQKLADADKQDAKTTLGTLDHLNDAQKQALTTQVEQAPDIATVNNVKQNAQNLN 1193

Qy  1028 ------LQSKLETL------DLGEHGAELL--NDMKRFRDELEQS 1058
Db  1194 NAMTNLNNALQDKTETLNSINFTDADQAKKDDYTNAVSHAEGILSKANGSNASQTVEQA 1253

Qy  1059 ATRSVTVLGQHQGVLKSNGEIN-------SEFKPSPGKALVQSF------NVN 1098
Db  1254 MQR---VNEAKQALNGNDNVQRAKDAAKQVITNANDLNQAQKDALKQQVDAAQTVANVN 1309

Qy  1099 ---RSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVD------MSHQKGEIPLG 1147
Db  1310 TIKQTAQDLNQAMTQLKQGIADKDQTKAN--GNFVNADTDKQNAYNNAVAHAEGII--- 1363

Qy  1148 RQRDPNDKTALTKSRLILDTV-TIGEL---HEI--ADKAKLVSDHKPDADQ 1192
Db  1364 -SGTPNANVDPQQVAQALQQVNQAKGDLNGNHNLQVAKDNANTAIDQLPNLNQPQKTALK 1422

Qy  1193 --------IKQ---LRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAV 1234
Db  1423 DQVSHAELVTGVNAIKQNADALNNAMGTLKQIQANSQVPQSVD--FT---QADQDKQ 1475

Qy  1235 KAFINAFKKEHHGVNLTTRTVLES-------------QSAELAKKLKNTLLSLDS 1277
Db  1476 QAYNNAANQAQQIANGTPTPVLAPDTVKAVTTMNQAKDALNGDEKLAQAKQDALANLDT 1535

Qy  1278 GESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYNLS----FSRTSGGLNVS 1333
Db  1536 LRDLN--------QPQRDALRNQINQAQALATVEQTKQNAQNVNTA 1573

Qy  1334 FGR-DGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWL------SAKHKISPDLRIGAV 1386
Db  1574 MGNLKQGIANKDTVKASENYHDADVDKQTAYTNAVSQAEGIINQTTNPTLNPDDITRALT 1633

Qy  1387 SGT-LQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIE------HQMKQGS 1437
Db  1634 QVTDAKNSLNGEAKLATEKQNAKDAVSGMTHLNDAQKQALKGQIDQSPEIATVNQVKQ- 1691
```

```
Qy  662  ALVIDNQLGLHHTNP--EPHEILDMGHLGLSLALQEGKLHYFDQLTKGWTGAESDCKQ---  716
              :  ::    ||    :    :   ::      : :
Db  290  AKAITNQT----SNPTMEPNTITQV-TSQVVTKEQALNGARNLAQAKTTAKNNLNNLTS   343

Qy  717  ------LKKGLDGAAYL-LKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGD     766
               ::   ||         |||    :    :  :
Db  344  INNAQKDALTRSIDGATTVAGVNQETAKATELNNAMHSLQNG---INDETQTKQ        394

Qy  767  ALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELK   826
              :|     :: :    |||:   :    :  ::  : |:
Db  395  TQKYLDAEPSKKSAYDQAVNAAKAILTKASGQNVDKAAVEQALQNVNSTKTALNGDAKLN   454

Qy  827  DIHVDHKQNLYALTHEGEVFHQPREAWQN----GAESSSWHKL------ALPQSESK     873
              :    :   :                            :|
Db  455  EAKAAAKQTLGTLTH---INNAQRTALDNEITQATNVEGVNTVKAKAQQLDGAMGQLETS   511

Qy  874  LKSLDMSHEHKPIATFEDGSQHQLK-----------AGGWHAYAAPERGPLAV-        915
              :: |||               :            |||
Db  512  IRDKDTTLQSONYQDADDAKFTAYSQAVNAAATILNKTAGGNTPKADVERAMQAVTQANT   571

Qy  916  GTSGSQTVFNRLMQGVKGKVIPGSGLTVK----LSAQ---TGGMTGAEG-------EV   957
              :                   |||                        :  ::
Db  572  ALNGIQNL-DRAKQAANTAITNASDLNTKQEEALKAQVTSAGRVSAANGVEHTATELNTA   630

Qy  958  ------RKVSSK--------FSERIRAYAFNPTMSTPRP------IKNAA          987
               : :: |            :   :                       :|
Db  631  MTALKRAIADKAETKASGNYVNADANKRQAYDEKVTA-AENIVSGTPTPTLTPADVTNAA   689

Qy  988  ------YATQHGWQGREGLKPLYEMQGALIKQLDAHVRHNAPQPDLQS              1030
               :::   :            : :         : :::
Db  690  TQVTNAKTQLNGNHNLEVAKQNANTAIDGLTSLNGPQKAKLKE---QVGQATTLPNVQT   745

Qy 1031  KLETLDLGEHGAELLND-MKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGK  1089
              :  ||  ||                  :  ::     | ::    :    :
Db  746  VRD--------NAQTLNTAMKGLRDSIANEATIKA---GQNYTDASQNKQTDYNSAVTAAK  795

Qy 1090  ALVQSFNVNRSGQDLSKSLQ-QAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGR  1148
              :  ::                                              |||
Db  796  AII--------GQTTSPSMNAQEINQAKDQVTAKQQALNG-------------------   827

Qy 1149  QRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDA---------DQIKQLR   1197
              :     :                                            :  ::
Db  828  --QENLRTAQTNAK-------QHLNGLSD---LTDAQKDAVKRQIEGATHVNEVTQAQ    873

Qy 1198  QQFD------TLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNL   1250
              :      :         : :
Db  874  NNADALNTAMTNLKNGIQDQNTIK---QGVNFTDADEAKRN---AYTNAVTQAEQILNK  926

Qy 1251  T-------TRTVLESQGSAE--------LAKKLKNTLLSLDSGESMSFSRSY        1287
               :            :                 |::
Db  927  AQGPNTSKDGVETALENVQRAKNELNGNQNVANAKTTAKNALNNLTSINNAQKEALKSQI  986

Qy 1288  GGGVSTVFVPTLSKKVPVVIPGAGITLDRAYNLSFSRTSGGLN----VSFGR        1336
              ::                                             |:||
Db  987  EGATTVAGVNQVS--------TTASELNTAMSNLQNGINDEAATKAAQKYTDADR      1033

Qy 1337  DGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWL----SAKHKISPDLRIGAAVSGTLQ-  1391
              ::                                     :|||
Db 1034  EKQTAYNDAVTAAKTLLDKTAGSNDNKAAVEQALQRVNTAKTALNGDERLNEAKNTAKQQ  1093

Qy 1392  ---GTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHOMKQGSKLTFSVDTS-   1446
                 :                                                  :|
Db 1094  VATMSHLTDAQKANLTSQIESGTTVAGVQIQANAGTLDQAM-NQLRQSIASKDATKSSE   1152

Qy 1447  ----ANLDLRAGIN----LNEDGSKPNGVTARVS----AGLSASANLAAG          1484
                  :                  |||                  :|||
Db 1153  DYQDANADLQNAYNDAVTNAEGIISATNNPEMNPDTINQKASQVNSAKSALNGDEKLAAA  1212

Qy 1485  SRERSTTSGQFGSTTSASNNRPTFLNG-VGAGANLTAALGVAHSSTHEGKPVGIFPAFTS  1543
              :|                                                :|
Db 1213  --KQTAKSDIGRLTDLNNAQRTAANAEVDQAPNLAAVTAAKNKATSLNTAMG------   1262
```

```
Qy 1438  KLTFSVDTSANLD-------------LRAGINLNEDGSKPNGVTARVSAGLSASANL    1481
              :   :: |                    |||
Db 1692  ------TATSLDQAMDQLSQAINDKDQILADGNYLNADPDKQNAYKQAV-AKAEALLNK  1743

Qy 1482  AAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAA--------LGVAHSSTHEGK  1533
              :||                                       |||
Db 1744  QSGTNE---VQAQVESITNEVNAAKQALNG---NDNLANAKQQAKQQLANLTHLNDAQKQ  1797

Qy 1534  PVGIFPAFTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSAT--  1591
              |:                                                    :|
Db 1798  ------SFESQITQAPLVTDVTTINQKAQTLDHAMELLRNSVADNQTTLASEDYHDATAQ  1851

Qy 1592  ----TKMLAALKEL-------DDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLK      1635
                  :|                    :|||
Db 1852  RQNDYNKAVTAANNIINQTTSPTMNPDDVNGA----TTQVNNTKVALDGDENLAAAK---  1904

Qy 1636  KLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMM  1695
              :||
Db 1905  ------QQANNRLDQLDHLNNAQKQLQSQTQSSDIAAVNGHKQTAESLNTA--MGNLI   1956

Qy 1696  NNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVGRE---------EV   1744
              :||                                               :  ::
Db 1957  N---AIADHQAVEQRGNFINADTDKQTAYNTAVNEAAAMINKQTGQNANQTEVEQAITKV  2013

Qy 1745  GVLFQDRN---NLRVKSVSYSQSVSKSEGFNTP-------ALLLGTSNSAAMSMERNIGT  1794
              :||                : :  :                  :|||
Db 2014  QTTIQALNGDHNLQVAKTNATQAIDALTSLNDPQKTALKDQVTAATLVTAVHQIEQNANT  2073

Qy 1795  IN  1796
              :|
Db 2074  LN  2075
```

```
Query Match      2.6%;  Score 242;  DB 23;  Length 2434;
Best Local Similarity  18.2%;  Pred. No. 5e-08;
Matches  311;  Conservative 238;  Mismatches 656;  Indels 500;  Gaps 73;

Qy  493  QNLSDNKSSEKLVDKI-KSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHF   551
              |||   :                                    :|||
Db   73  QNLAQAKTN-ATNTINNAHDLNQKQKDALKTQVNNAQRVS----DANNVQHTATELNS    125

Qy  552  A-DAHQGLLHGKSELEAQSVAISHGRLVVADSE------------------------    583
              :                                    :|||
Db  126  AMTALKAAIADKERTKAS-----GNYVNADQEKRQAYDSKVTNAENIISGTPNATLTVN   179

Qy  584  ----GKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNAL---   634
               ::  |||                              :
Db  180  DVNSAASQVNAAKTALNGDNN--LRVAKEHA-------NNTIDGLAQLNNAQKAKLKEQ  229

Qy  635  ---------VKNNFRQQHACPLG----NDHQFHPGWNLTD-------------------  661
                    :|                                 :|
Db  230  VQSATTLDGVQTVKNSSQTLNTAMKGLRDSIANEATIKAGQNYTDASPNNRNEYDSAVTA  289
```

```
Qy  662 ALVIDNQLGLHITNP-EPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQ---  716
            | | :| ||    :   :  ||  : :| : :   |  | : |:| ||||  :||
Db  290 AKAIINQT---SNPTMEPNTTQV--TSQVTTKEQALNGARNLAQAKTAKNNLNNLTS    343

Qy  717 ------LKKGLDGAAYL--LKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGD   766
              : ::| ::|||    | ::  ::   ::::|    : : :|       |
Db  344 INNAQKDALTRSIDGATTVAGVNQETAKATELNNAMHSLQNG---INDETQTKQ       394

Qy  767 ALQGLNKDDKDKQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELK 826
            | || :::|  |  |  :||  : | ||     :|  :::|  | :| | :
Db  395 TQKYLDAEPSKKSAYDQAVNAAKAILTKASGQNVDKAAVEQALQNVNSTKTALNGDAKLN   454

Qy  827 DIHVDHKQNLYALTHEGEVFHQPREAWON----GAESSWHKL----ALPQSESK       873
            :: :||| |  : :|| |::        :|   :|:  :        | |:||
Db  455 EAKAAAKQTLGTLTH---INNAQRTALDNEITQATNVEGVNTVKAKAQQLDGAMGQLETS   511

Qy  874 LKSLDMSHEHKPIATFEDGSQHOLK-----------AGGWHAYAAPERGPLAV---      915
            | :|   | :: : |  :|   ||            ||  || |
Db  512 IRDKDTTLQSQNYQDADDAKRfAYSQAVNAAATILNKTAGGNTPKADVERAMQAVTQANT   571

Qy  916 GTSGSQTVFNRLMQGVKGKVIPGSGLTVK-----LSAQ--TGGMTGAEG------      957
            ::   ||  ||:|        |  ||:::    : : |   || | ||
Db  572 ALNGIQNL-DRAKQAANTAITNASDLNTKQKEALKAQVTSAGRVSAANGVEHTATELNTA   630

Qy  958 -----RKVSSK------------FSERIRAYAFNPTMSTPRP-----IKNAA         987
             : : : ::             | :||:|  ||:|:|| ||     |: ||
Db  631 MTALKRAIADKAETKASGNYVNADANKRQAYDEKVTA-AENIVSGTPTPTLTPADVTNAA   689

Qy  988 ---------------YATOHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDLQS   1030
                           :   |   ::  : | :|  |:|||  : :  :  :::::
Db  690 TQVTNAKTQLNGNHNLEVAKQNANTAIDGLTSLNGPQKAKLKE---QVGQATTLPNVQT   745

Qy 1031 KLETLDLGEHGAELLND-MKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGK  1089
            : | || ||  :| :: ::   ||:  :|  ::   :||::|::| | |:::  :|||
Db  746 VRD----NAQTLNTAMKGLRDSIANEATIKA--GQNYTDASQNKQTDYNSAVTAAK      795

Qy 1090 ALVQSFNVNRSGQDLSKSLQ-QAVHATPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGR   1148
            | :::||:|| |  | ::|| :|:   |:|:::|:|:|||||||  |  ||| |:| ||
Db  796 AII----GQTTSPSMNAQEINQAKDQVTAKQQALNG                         827

Qy 1149 QRDPNDKTALTKSRLLDTVTIGELHELADRAKLVSDHKPDA--------DQIKQLR     1197
            ::     | |  :||| |   :| |: |:|||| | :|||         |:|:|
Db  828 --QENLRTAQTNAK--------QHLNGLSD---LTDAQKDAVKRQIEGATHVNEVTQAQ   873

Qy 1198 QQFD-------TLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNL  1250
            ::::        |::::::|:|  :: | | ::|||||||:|||||| ||:||:|| |||
Db  874 NNADALNTAMTNLKNGIQDQNTIK---QGVNFTDADEAKRN--AYTNAVTQAEQILNK     926

Qy 1251 T------TRTVLESQGSAE-----------LAKKLKNTLLSLDSGESMSFSRSY       1287
            |      | ||||| ::|:           ||||||||| | |:||:|:||||
Db  927 AQGPNTSKDGVETALENVQRAKNELNGNQNVANAKTTAKNALNNLTSNNAQKEALKSQI   986

Qy 1288 GGGVSVTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLN------VSFGR     1336
            :| ||: |:  |::|||    ||| ||||||||||||:||:||:|       |:|:
Db  987 EGATTVAGVNQVS----------TTASELNTAMSNLQNGINDERAATKAAQKYTDADR  1033

Qy 1337 DGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWL----SAKIKISPDLRIGAVSGTLQ-  1391
            :||||||| | |||| || ||||||:|||:|:||    |||:||||||||||:||||
Db 1034 EKQTAYNDAVTAAKTLLDKTAGSNDNKAAVEQALQRVNTAKTALNGDERLNEAKNTAKQQ 1093

Qy 1392 ----GTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTS- 1446
                ||||||||||||||||||||||||||||||||||:|||:|:||:||||||:|||
Db 1094 VATMSHLTDAQKANLTSQIESGTTVAGVQGIQANAGTLDQAM-NQLRQSIASKDATKSSE 1152

Qy 1447 ----ANLDLRAGIN-------LNEDGSKPNGVTARVS-------AGLSASANLAAG     1484
                |||:|:||:|       |||:|:||||||||:|       |||||:||||||
Db 1153 DYQDANADLQNAYNDAVTNAEGIISATNNPEMNPDTINQKASQVNSAKSALNGDEKLAAA 1212

Qy 1485 SRERSTTSGQFGSTTSASNNRPTFLNG-VGAGANLTAALGVAHSSTHEGKPVGIFPAFTS 1543
            :::||||||:|||||||||||||:|||  |||||||||||||||||||||||||||||||
Db 1213 --KQTAKSDIGRLTDLNNAQRTAANAEVDQAPNLAAVTAAKNKATSLNTAMG------   1262

Qy 1544 TNVSAALALDNRTSQSISL-----ELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAAL 1598
            :|:||||||||||||||||     ||||||||||||||||||||||||| :|||||||||
Db 1263 -NLKHALAEKDNTKRSVNYTDADQPKQQAYDTAVTQAEAITNANGSNANE---TQVQAAL 1318

Qy 1599 KELDDAKPAEQLH----ILQQHFSAKDVVGDERYEAVRNLKLVIRQQ----------A  1643
            :|| :||  |::|    || ||:||||  ||:|| | |:|| :  :|           |
Db 1319 NQLNQAK--NDLNGDNKVAQAKESAKRALAS--YSNLNNAQSTAATSQIDNATTVAGVTA 1374

Qy 1644 ADSHSMELGSA-------------SHSTTYNNLSRINNDGIVELLH-              1676
            |:   | | :|             |:|:|:|:  ::||| |: :|
Db 1375 AQNTANELNTAMGQLONGINDQNTVKQQVNFTDADQKKDAYTN-AVTNAQGILDKAHGQ  1433

Qy 1677 ---KHFDAAL------------PASSAK-RLGEMMN-NDPALKDIIKQLQSTP       1712
               |: | ||            || |:| :| ::|: ||  |:| : :|:|:|
Db 1434 NMTKAQVEAALNQVTTAKNALNGDANVRQAKSDAKANLGTLTHLNNAQKQDLTSQIEGAT 1493

Qy 1713 FSSASVSMELK----DGLREQTEKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQS---V 1765
            ::|||:|: ||    ||  ||::|| :||| | :  || |:| : :  |:: :|:|   |
Db 1494 TVNGVNGVKTKAQDLDGAMQRLQSAIANKDQTKASENVIDADPTKKTAFDNAITQAESYL 1553

Qy 1766 SKSEGFNTPALLGTSNSAAMSMERNI-GTINFKYGQDQ-----------NTPRRFTLEG  1813
            :|:  : :: | |:||:|| :| ||: |: | :: |:|           ||| :| | |
Db 1554 NKDHGANKDKQAVEQAIQSVTSTENALNGDANLQRAKTEAIQAIDNLTHLNTPQKTALKQ 1613

Qy 1814 GIAQANPQVASALTDLKKEGLEMKS                                   1838
            |: :||| |:| |||||::  |: |
Db 1614 QVNAA--QRVSGVTDLKNSATSLNN                                   1636
```

Query Match   2.6%; Score 242; DB 23; Length 2434;
Best Local Similarity   18.2%; Pred. No. 5e-08;
Matches 311; Conservative 238; Mismatches 656; Indels 500; Gaps 73;

```
Qy  493 QNLSDNKSSEKLVDKI-KSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHF  551
            |||: |:| | ::| :   :|:  : ||: | :  :: |:::  : |  :|:   :::
Db   73 QNLAQAKTN--ATNTINNAHDLNQKQKDALKTQVNNAQRVS----DANNVQHTATELNS   125

Qy  552 A-DAHQGLLHGKSELEAQSVAISHGRLVVADSE------                      583
            | || : :: |:| :|: |:| : | :::||
Db  126 AMTALKAAIADKERTKAS----GNYVNADQEKRQAYDSKVTNAENIISGTPNATLTVN    179

Qy  584 ----GKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNAL-      634
                |:| ||||  ||:||:|:  || :  |  :: :|||:|| :: :| ||:|
Db  180 DVNSAASQVNAAKTALNGDNN--LRVAKEHA------NNTIDGLAQLNNAQKAKLKEQ   229

Qy  635 ---------VKNNFRQQHACPLG---NDHQFHPGWNLTD---------             661
                     ||:| ||: :| :|   ||:| : | | ||
Db  230 VQSATTLDGVQTVKNSSQTLNTAMKGLRDSIANEATIKAGQNYTDASPNNRNEYDSAVTA 289
```

```
Db   290  AKAIINQT----SNPTMEPNTIQV--TSQVTTKEQALNGARNLAQAKTAKNNLNNLTS    343
          :|  |||                    |:| :                   |:: |||
Qy   717  ------LKKGLDGAAYL--LKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGD    766

Db   344  INNAQKDALTRSIDGATTVAGVNQETAKATELNNAMHSLQNG--------------INDETQTKQ  394

Qy   767  ALQGLNKDKDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELK  826

Db   395  TQKYLDAEPSKKSAYDQAVNAAKAILTKASGQNVDKAAVEQALQNVNSTKTALNGDAKLN  454

Qy   827  DIHVDHKQNLYALTHEGEVFHQPREAWQN-------GAESSSWHKL--ALPQSESK  873

Db   455  EAKAAAKQTLGTLTH--INNAQRTALDNEITQATNVEGVNTVKAKAQQLDGAMGQLETS  511

Qy   874  LKSLDMSHEHKPIATFEDGSQHQLK--------------AGWHAYAAPERGPLAV--  915

Db   512  IRDKDTTLQSQNYQDADDAKRTAYSQAVNAAATILNKTAGGNTPKADVERAMQAVTQANT  571

Qy   916  GTSGSQTVFNRLMQGVKGKVIPGSGLTVK-------LSAQ--TGGMTGAEG------  957

Db   572  ALNGIQNL-DRAKQAANTAITNASDLNTKQKEALKAQVTSAGRVSAANGVEHTATELNTA  630

Qy   958  -----RKVSSK-----------FSERIRAYAFNPTMSTPRP------IKNAA  987

Db   631  MTALKRAIADKRAETKASGNYVNADANKRQAYDEKVTA-AENIVSGTPTPTLITPADVTNAA  689

Qy   988  ----------YATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDLQS  1030

Db   690  TQVTNAKTQLNGNHNLEVAKQNANTAIDGLTSLNGPQKAKLKE---QVGQATTLPNVQT  745

Qy  1031  KLETLDLGEHGAELLND-MKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGK  1089

Db   746  VRD------NAQTLNTAMKGLRDSIANEATIKA--GQNYTDASQNKQTDYNSAVTAAK  795

Qy  1090  ALVQSFNVNRSGQDLSKSLQ-QAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGR  1148

Db   796  AII------GQTTSPSMNAQEINQAKDQVTAKQQALNG-  827

Qy  1149  QRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDA--------DQIKQLR  1197

Db   828  --QENLRTAQTNAK-------QHLNGLSD--LTDAQKDAVKRQIEGATHVNEVTQAQ  873

Qy  1198  QQFD-----------TLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNL  1250

Db   874  NNADALNTAMTNLKNGIQDQNTIK---QGVNFTDADEAKRN--AYTNAVTQAEQILNK  926

Qy  1251  T-----------TRTVLESQGSAE----------LAKKLKNTLLSLDSGESMSFSRSY  1287

Db   927  AQGPNTSKDGVETALENVQRAKNELNGNQNVANAKTTAKNALNNLTSINNAQKEALKSQI  986

Qy  1288  GGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLN-------VSFGR  1336

Db   987  EGATTVAGVNQVS---------TTASELNTAMSNLQNGINDEAATKAAQKYTDADR  1033

Qy  1337  DGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWL---SAKHKISPDLRIGAAVSGTLQ-  1391

Db  1034  EKQTAYNDAVTAAKTLLDKTAGSNDNKAAVEQALQRVNTAKTALNGDERLNEAKNTAKQQ  1093

Qy  1392  ------GTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTS-  1446

Db  1094  VATMSHLTDAQKANLTSQIESGTTVAGVQIQANAGTLDQAM-NQLRQSIASKDATKSSE  1152

Qy  1447  -----ANLDLRAGIN---------LNEDGSKPNGVTARVS--------AGLSASANLAAG  1484

Db  1153  DYQDANADLQNAYNDAVTNAEGIISATNNPEMNPDTINQKASQVNSAKSALNGDEKLAAA  1212

Qy  1485  SRERSTTSQOFGSTTSASNRPTFLNG-VGAGANLTAALGVAHSSTHEGKPVGIFPAFTS  1543

Db  1213  --KQTAKSDIGRLTDLNNAQRTAANAEVDQAPNLAAVTAAKNKATSLNTAMG------  1262

Qy  1544  TNVSAALALDNRTSQSISL----ELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAAL  1598

Db  1263  -NLKHALAEKDNTKRSVNYTDADQPKQQAYDTAVTQAEAITNANGSNANE---TQYQAAL  1318
```

```
Db  1263  -NLKHALAEKDNTKRSVNYTDADQPKQQAYDFAVTQAEAITNANGSNANE---TQYQAAL  1318

Qy  1599  KELDDAKPAEQLH------ILQQHFSAKDVVGDERYEAVRNLKKLVIRQQ----------A  1643

Db  1319  NQLNQAK--NDLNGDNKVAQAKESAKRALAS--YSNLNNAQSTAATSOIDNATTVAGVTA  1374

Qy  1644  ADSHSMELGSA----------SHSTTYNNLSRINNDGIVELLH--  1676

Db  1375  AQNTANELNTAMGQLQNGINDQNTVKQQVNFTDADQGKKDAYTN-AVTNAQGILDKAHGQ  1433

Qy  1677  ---KHFDAAL----------PASSAK-RLGEMMN-NDPALKDIIKQLQSTP  1712

Db  1434  NMTKAQVEAALNQVTTAKNALNGDANVRQAKSDAKANLGTLTHLNNAQKQDLTSQIEGAT  1493

Qy  1713  FSSASVSMELK----DGLREQTEKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQS---V  1765

Db  1494  TVNGVNGVKTKAQDLDGAMQRLQSAIANKDQTKASENYIDADPTKKTAFDNAITQAESYL  1553

Qy  1766  SKSEGFNTPALLGTSNSAAMSMERNI-GTINFKYGQDQ----------NTPRRFTLEG  1813

Db  1554  NKDHGANKDKQAVEQAIQSVTSENALNGDANLQRAKTEAIQAIDNLTHLNTPQKTALKQ  1613

Qy  1814  GIAQANPQVASALTDLKKEGLEMKS  1838

Db  1614  QVNAA--QRVSGVTDLKNSATSLNN  1636
```

RESULT 14
US-60-257-931-3392
; Sequence 3392, Application US/60257931
; GENERAL INFORMATION:
;   APPLICANT: Haselbeck, R
;   APPLICANT: Ohlsen, KL
;   APPLICANT: Zyskind, JW
;   APPLICANT: Trawick, JD
;   APPLICANT: Wall, D
;   TITLE OF INVENTION: Genes identified as essential in Staphylococcus aureus, Salmonel
;   TITLE OF INVENTION: typhimurium, Klebsiella pneumoniae and Pseudomonas aeruginosa
;   FILE REFERENCE: ELITRA.017PR4
;   CURRENT APPLICATION NUMBER: US/60/257,931
;   CURRENT FILING DATE: 2000-12-22
;   NUMBER OF SEQ ID NOS: 3592
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3392
;   LENGTH: 2434
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-60-257-931-3392

```
Query Match    2.6%;  Score 242;  DB 23;  Length 2434;
Best Local Similarity  18.2%;  Pred. No. 5e-08;
Matches  311;  Conservative 238;  Mismatches 656;  Indels 500;  Gaps  73;

Qy   493  QNLSDNKSSEKLVDKI-KSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHF  551

Db    73  QNLAQAKTN--ATNTINNAHDLNQKQDALKTQVNNAQRVS---------DANNVQHTATELNS  125

Qy   552  A-DAHQGLLHGKSELEAQSVAISHGRLVVADSE----------  583

Db   126  AMTALKAAIADKERTKAS-----GNYVNADQEKRQAYDSKVTNAENIISGTPNATLTVN  179

Qy   584  -----GKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNAL---  634

Db   180  DVNSAASQVNAAKTALNGDNN--LRVAKEHA---------NNTIDGLAQLNNAQKAKLKEQ  229

Qy   635  ------VKNNFRQQHACPLG-----NDHQFHPGWNLTD---------  661

Db   230  VQSATTLDGVQTVKNSSQTLNTAMKGLRDSIANEATIKAGQNYTDASPNNREYDSAVTA  289

Qy   662  ALVIDNQLGLHHTNP--EPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQ---  716
```

```
Qy   1599  KELDDAKPAEQLH------ILQQHFSAKDVVGDERYEAVRNLKKLVIRQQ-----------A 1643
                :|:.  |||| .:   :||||      .||| :: |:|:
Db   1319  NQLNQAK--NDLNGDNKVAQAKESAKRALAS--YSNLNNAQSTAATSQIDNATTVAGVTA   1374

Qy   1644  ADSHSMELGSA--------------------SHSTTYNNLSRINNDGIVELLH--       1676
                |:.| :|:|                    | ||:    :|| ||  || |
Db   1375  AQNTANELNTAMGQLQNGINDQNTVKQQVNFTDADQGKKDAYTN-AVTNAQGILDKAHGQ  1433

Qy   1677  ---KHFDAAL--------------------PASSAK-RLGEMNN-NDPALKDIIKQLQSTP  1712
                 |||        |||   |         :|| ||     :| ||:
Db   1434  NMTKAQVEARLNQVTTAKNALNGDANVRQAKSDAKANLGTLHLNNAQKQDLTSQIEGAT   1493

Qy   1713  FSSASVSMELK---DGLREQTEKAILDGKVGREEVGLFQDRNNLRVKSVSVSQS----V  1765
                || ||:|       |    :|    ||  |    :.            |||:
Db   1494  TVNGVNGVKTKAQLDGAMQRLQSAIANKDQTKASENYIDADPTKKTAFDNAITQAESYL  1553

Qy   1766  SKSEGFNTPALLLGTSNSAAMSMERNI-GTINFKYGQDQ----------NTPRRFTLEG   1813
                | | ||| |||:||   | ::  | :   |  | || |         |||: :|:
Db   1554  NKDHGANKDQAVEQAIQSVTSTENALNGDANLQRAKTEAIQAIDNLTHLNTPQKTALKQ  1613

Qy   1814  GIAQANPQVASALTDLKKEGLEMKS 1838
                 |:|    | :|||  || :
Db   1614  QVNAA--QRVSGVTDLKNSATSLNN 1636
```

RESULT 15
US-60-269-308-4413
; Sequence 4413, Application US/60269308
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, RW
; APPLICANT: Ohlsen, KL
; APPLICANT: Zyskind, JW
; APPLICANT: Trawick, JD
; APPLICANT: Wall, D
; TITLE OF INVENTION: Identification of Essential Genes in Staphylococcus aureus, Pseud
; TITLE OF INVENTION: aeruginosa, Klebsiella pneumoniae, Salmonella typhimurium, and E
; TITLE OF INVENTION: faecalis
; FILE REFERENCE: ELITRA.017PR5
; CURRENT APPLICATION NUMBER: US/60/269,308
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 4774
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4413
; LENGTH: 2434
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-60-269-308-4413

Query Match   2.6%; Score 242; DB 23; Length 2434;
Best Local Similarity 18.2%; Pred. No. 5e-08;
Matches 311; Conservative 238; Mismatches 656; Indels 500; Gaps 73;

```
Qy    493  QNLSDNKSSSEKLVDKI-KSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHF  551
                ||||:    ||: ||                .| |:    .|.::|  ||
Db     73  QNLAQAKTN-ATNTINNAHDLNQKQKDALKTQVNNAQRVS----DANNVQHTATELNS    125

Qy    552  A-DAHQGLLHGKSELEAQSVAISHGRLVVADSE------------------GNYVNADQ   583
               | |:||:|| |  :||    :|.:||  |              :|  |  :||
Db    126  AMTALKAAIADKERTKAS----GNYVNADQEKRQAYDSKVTNAENIISGTPNATLTVN   179

Qy    584  ---GKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNAL----   634
                  |.  :| |::|:|| ||  | .: |  :: ::|  .|
Db    180  DVNSAASQVNAAKTALNGDNN--LRVAKEHA------NNTIDGLAQLNNAQKAKLKEQ   229

Qy    635  --------VKNNFRQQHACPLG------NDHQFHPGWNLTD---------          661
                    |:|::  |  |||        || :| |::|||
Db    230  VQSATTLDGVQTVKNSSQTLNTAMKGLRDSIANEATIKAGQNYTDASPNNRNEYDSAVTA  289

Qy    662  ALVIDNQLGLHHTNP--EPHEILDMGHLGSLALQEKLHYFDQLTKGWTGAESDCKQ---  716
                ||: |   .|:   .   | |   |  |: |  |:::||  ::|.:::   ||
Db    290  -NLKHALAEKDNTKRSVNYTDADQPKQQAYDTAVTQAEAITNANGSNANE---TQVQAAL  343
```

```
Qy    717  ------LKKGLDGAAYL--LKDGEVKRLNLINQSTSSIKHGTENVFSLPHVRNKPEPGD  766
                  :|: . :|||      .||| :: |:|:
Db    344  INNAQKDALTRSIDGATTVAGVNQETAKATELNNAMHSLQNG-------INDETQTKQ   394

Qy    767  ALQGLNKDKRAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTLSREGISGELK  826
                |::|::  || ::||| :|  : || ::|  |:  | :
Db    395  TQKYLDAEPSKKSAYDQAVNAAKAILTKASGQNVDKAAVEQALQNVNSTKTALNGDAKLN  454

Qy    827  DIHVDHKQNLYALTHEGEVFHQPREAWQN-------GAESSSWHKL----ALPQSESK    873
                |:|  ||  |  : :  : ||    |        || |   |   |
Db    455  EAKAAAKQTLGTLTH--INNAQRTALDNEITQATNVEGVNTVKAKAQQLDGAMGQLETS  511

Qy    874  LKSLDMSHEHKPIATFEDGSQHQLK------------AGGWHAYAAPERGPLAV----  915
                |: |:     |: ||||:|  ||
Db    512  IRDKDTTLQSNYQDADDAKRTAYSQAVNAAATILNKTAGGNTPKADVERAMQAVTQANT  571

Qy    916  GTSGSQTVFNRLMQCVKGKVIPGSGLTVK---LSAQ---TGGMTGAEG----------  957
                |:|||||:|  |:: |   ||| :|    | |    |||| ||
Db    572  ALNGIQNL-DRAKQAANTAITNASDLNTKQKEALKAQVTSAGRVSAANGVEHTATELNTA 630

Qy    958  ----RKVSSK------------------FSERIRAYAFNPTMSTPRP------IKNAA  987
                .:::       :    :||  |  :   ::| |  || :   |  :||||
Db    631  MTALKRAIADKAETKASGNYVNADANKRQAYDEKVTA-AENIVSGTPTPTLTPADVTNAA  689

Qy    988  --------------------YATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDLQS 1030
                                   :| |:.  | :| :|   : |::|   | |:
Db    690  TQVTNAKTQLNGNHNLEVAKQNANTAIDGLTSLNGPQKAKLKE----QVGQATTLPNVQT  745

Qy   1031  KLETLDLGEHGAELLND-MKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGK 1089
                |:  | | |  | |  |  ::|  .|  |  ||:   |   |: ::  |
Db    746  VRD---NAQTLNTAMKGLRDSIANEATIKA--GQNYTDASQNKQTDYNSAVTAAK    795

Qy   1090  ALVQSFNVNRSGQDLSKSLQ-QAVHATPPSAESKIQSMLGHFVSAGVDMSHQKGEIPLGR 1148
                |:::     |   |      | .:|::||   .:.:|::
Db    796  AII----GQTTSPSMNAQEINQAKDQVTAKQQALNG-----------------------  827

Qy   1149  QRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDA----------DQIKQLR 1197
                .| .||  |: ||:|  :|  | :.:   : :  :  |          :|:| |
Db    828  --QENLRTAQTNAK------QHLNGLSD---LTDAQKDAVKRQIEGATHVNEVTQAQ   873

Qy   1198  QQFD------TLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNL 1250
                || |       |::||:|:  |: | ::|:   :||  |  |:|  :|
Db    874  NNADALNTAMTNLKNGIQDQNTIK------QGVNFTDADEAKRN--AYTNAVTQAEQILNK  926

Qy   1251  T-------TRTVLESOGSAE---------------LAKKLKNTLLSLDSGESMSFSRSY 1287
                |        | |  | ::                |:|:|  |::|  :  .:: ::|
Db    927  AQGPNTSKDGVETALENVQRAKNELNGNQNVANAKTTAKNALNNLTSINNAQKEALKSQI  986

Qy   1288  GGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLN----------VSFGR 1336
                || |:| :::::  |:|:    || :|  :|  ||: ::| :             | |
Db    987  EGATTVAGVNQVS------TTASELNTAMSNLQNGINDEAATKAAQKYTDADR   1033

Qy   1337  DGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWL---SAKHKISPDLRIGAVSGTLQ-  1391
                || |.| |  |:| |   || :||||: ||:|     || :|||  :|  |.|| |
Db   1034  EKQTAYNDAVTAAKTLLDKTAGSNDNKAAVEQALQRVNTAKTALNGDERLNEAKNTAKQQ 1093

Qy   1392  ----GTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTS- 1446
                    ||: |: :    |:|   || |  :| | ||   :|| : |:: |: :|:  |
Db   1094  VATMSHLTDAQKANLTSQIESCTTVAGVQGIQANAGTLDQAM-NQLRQSIASKDATKSSE 1152

Qy   1447  ---ANLDLRAGIN-------------LNEDGSKPNGVTARVS-----AGLSASANLAAG  1484
                   ||:| |: |                |::| |:|| ||::         |:|:|| |
Db   1153  DYQDANADLQNAYNDAVTNAEGIISATNNPEMNPDTINQKASQVNSAKSALNGDEKLAAA 1212

Qy   1485  SRERSTTSGQFGSTTSASNNRPTFLNG-VGAGANLTALGVAHSSTHEGKPVGIFPAFTS 1543
                | |:| | |  ||:| ||  ||: ||  | |  |: |   | |   ::|   |    |:
Db   1213  --KQTAKSDIGRLTDLNNAQRTAANAEVDQAPNLAAVTAAKNKATSLNTAMG-       1262

Qy   1544  TNVSAALALDNRTSQSISL----ELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAAL 1598
                ||: |:  |:: | | |      .||  .|:. | :  |:|:|| |:|:||  :|:|||
Db   1263  -NLKHALAEKDNTKRSVNYTDADQPKQQAYDTAVTQAEAITNANGSNANE---TQVQAAL 1318
```

```
Qy   1599  KELDDAKPAEQLH----ILQQHFSAKDVVGDERYEAVRNLKKLVIRQQ----------A 1643
                :|: || :  |  ||| : |:: |:  |                        |
Db   1319  NQLNQAK--NDLNGDNKVAQAKESAKRALAS--YSNLNNAQSTAATSQIDNATTVAGVTA 1374

Qy   1644  ADSHSMELGSA---------------------SHSTTYNNLSRINNDGIVELLH-- 1676
                |:: || :|                     | : | || || ::  |
Db   1375  AQNTANELNTAMGQLQNGINDQNTVKQQVNFTDADQGKKDAYTN-AVTNAQGILDKAHGQ 1433

Qy   1677  ---KHFDAAL----------PASSAK-RLGEMMN-NDPALKDIIKQLQSTP 1712
                :||              || || |: |:  :|: |::
Db   1434  NMTKAQVEAALNQVTTAKNALNGDANVRQAKSDAKANLGTLTHLNNAQKQDLTSQIEGAT 1493

Qy   1713  FSSASVSMELK----DGLREQTEKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQS---V 1765
                 :: | || :       |  ::  -  :|:|: |
Db   1494  TVNGVNGVKTKAQDLDGAMQRLQSAIANKDQTKASENYIDADPTKKTAFDNAITQAESYL 1553

Qy   1766  SKSEGFNTPALLLGTSNSAAMSMERNI-GTINFKYGQDQ----------NTPRRFTLEG 1813
                :| |||  : |  |   |: :                   ||:::: |:
Db   1554  NKDHGANKDKQAVEQAIQSVTSTENALNGDANLQRAKTEAIQAIDNLTHLNTPQKTALKQ 1613

Qy   1814  GIAQANPQVASALTDLKKEGLEMKS 1838
                :  | |  |:||||    :  :
Db   1614  QVNAA--QRVSGVTDLKNSATSLNN 1636
```

Search completed: June 5, 2001, 18:20:30
Job time: 356 sec

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    June  5, 2001, 18:13:54 ; Search time 26.26 Seconds
                                      (without alignments)
                                1344.614 Million cell updates/sec

Title:          US-09-596-784-2
Perfect score:  9448
Sequence:       1 MELKSLGTEHKAAVHTAAHN.........NPQVASALTDLKKEGLEMKS 1838
Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters:      185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued_Patents_AA:*
      1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
      2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
      3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
      4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
      5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
      6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

   Pred. No. is the number of results predicted by chance to have a
   score greater than or equal to the score of the result being printed,
   and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	209.5	2.2	1912	1	US-08-409-995-4	Sequence 4, Appli
2	209.5	2.2	1912	1	US-08-685-467-4	Sequence 4, Appli
3	203.5	2.2	1780	3	US-08-769-309A-5	Sequence 5, Appli
4	203.5	2.1	1780	3	US-08-994-570-5	Sequence 5, Appli
5	201	2.1	2353	4	US-09-377-155-33	Sequence 33, Appl
6	201	2.1	2353	4	US-08-913-942-4	Sequence 4, Appli
7	190.5	2.0	1536	1	US-08-038-682-2	Sequence 2, Appli
8	190.5	2.0	1536	2	US-08-302-832-2	Sequence 2, Appli
9	190.5	2.0	1536	2	US-08-530-198-2	Sequence 2, Appli
10	190.5	2.0	1536	2	US-08-469-880-2	Sequence 2, Appli
11	190.5	2.0	1536	2	US-08-728-470-2	Sequence 2, Appli
12	190.5	2.0	1536	2	US-08-617-697-2	Sequence 2, Appli
13	187	2.0	1600	2	US-08-617-697-10	Sequence 10, Appl
14	186.5	2.0	2482	2	US-08-328-254-6	Sequence 6, Appli
15	184.5	2.0	3248	5	PCT-US95-16216-1	Sequence 1, Appli
16	184.5	1.9	3248	2	US-08-617-697-9	Sequence 9, Appli
17	183.5	1.9	1599	2	US-08-353-700-1	Sequence 1, Appli
18	183	1.9	1477	1	US-08-038-682-4	Sequence 4, Appli
19	183	1.9	1477	2	US-08-530-832-4	Sequence 4, Appli
20	183	1.9	1477	2	US-08-530-198-4	Sequence 4, Appli
21	183	1.9	1477	2	US-08-469-880-4	Sequence 4, Appli
22	183	1.9	1477	2	US-08-728-470-4	Sequence 4, Appli
23	183	1.9	1477	2	US-08-617-697-4	Sequence 4, Appli
24	181	1.9	1529	2	US-08-728-470-10	Sequence 10, Appl
25	177	1.9	1338	2	US-08-728-470-9	Sequence 9, Appli
26	174	1.8	2842	1	US-07-741-940-2	Sequence 2, Appli
27	174	1.8	2842	1	US-08-289-548A-7	Sequence 7, Appli
28	174	1.8	2842	1	US-08-452-654-7	Sequence 7, Appli
29	174	1.8	2843	1	US-08-452-655B-2	Sequence 2, Appli
30	174	1.8	2843	1	US-08-452-655B-7	Sequence 7, Appli
31	174	1.8	2843	4	US-08-450-582-2	Sequence 2, Appli
32	174	1.8	2843	4	US-08-450-582-7	Sequence 7, Appli
33	174	1.8	2973	2	US-08-821-355A-7	Sequence 7, Appli
34	174	1.8	2973	2	US-09-003-687A-7	Sequence 7, Appli
35	170.5	1.8	2973	4	US-09-136-605-7	Sequence 5, Appli
36	170	1.8	3969	4	US-07-741-940-2	Sequence 2, Appli
37	170	1.8	2843	1	US-08-061-376-5	Sequence 2, Appli
38	170	1.8	2843	1	US-08-289-548A-2	Sequence 2, Appli
39	170	1.8	2843	1	US-08-452-654-2	Sequence 2, Appli
40	170	1.8	2843	2	US-08-370-235A-2	Sequence 2, Appli
41	169.5	1.8	1312	2	US-08-592-126-148	Sequence 148, App
42	169.5	1.8	1312	2	US-08-687-080-51	Sequence 51, Appl
43	169	1.8	1702	5	PCT-US95-10661A-5	Sequence 5, Appli
44	167	1.8	885	2	US-08-53-306A-4	Sequence 4, Appli
45	167	1.8	885	2	US-08-742-923A-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT 1
US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5646259
; GENERAL INFORMATION:
;   APPLICANT: Barenkamp, Stephen J.
;   APPLICANT: St. Geme III, Joseph W.
;   TITLE OF INVENTION:  Haemophilus Adhesion Proteins
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;     STREET:  Four Embarcadero Center, Suite 3400
;     CITY:  San Francisco
;     STATE:  CA
;     COUNTRY:  USA
;     ZIP:  94111-4187
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0,  Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/409,995
;     FILING DATE:  24-MAR-1995
;     CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Silva, Robin M.
;     REGISTRATION NUMBER:  38,304
;     REFERENCE/DOCKET NUMBER:  A-61053/RFT
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (415) 781-1989
;     TELEFAX:  (415) 398-3249
;     TELEX:  910 277299
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH:  1912 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  unknown
US-08-409-995-4

Query Match             2.2%;  Score 209.5;  DB 1;  Length 1912;
  Best Local Similarity  18.9%;  Pred. No. 8.1e-06;
  Matches  346;  Conservative  226;  Mismatches  653;  Indels  609;  Gaps  85;

Qy    236  SEFQQLHQQRLARERENPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQS  295

Db     20  SELTRTHTKRL-RNRGDPV------LATLLFATVQANATDEDEE-LDPVVRTAPVLSFHS   71
```

```
Db   946 GKLFTGKDLKDANNGATVSEDDGKD-----TGTGLVTAKTV---IDAVNK-SGWRVTG 994
Qy  1246 HGVNLTRTVLESOGSAELAKKLKNTLLSLDSGESMSFSRSYGGVSTVFVPTLSKKVPV 1305
Db   995 EGATAETGATAVNAGNAE------TVTSGTSVNFK--NGNATTATV----------- 1032
Qy  1306 PVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMP----YMTGKK 1360
Db  1033 ------------SKDNGNIVKY-DVNVGDGLKIGDDKKIVADTTLTVTGGK 1072
Qy  1361 TSAGNASDWLSAKHKISPDLRIGAAVSGTLQCTLQNSLKFFLTEDELPGFIHGLTHGTLT 1420
Db  1073 VSVPAGANSVNNNKKL--------VNAEGLATALNNLSWTAKADK-YADGESEGE- 1118
Qy  1421 PAELLQKGIEHQMKQGSKLTFSVDTSANLLDRAGINLNEDGSK--------PNGVTARVSA 1473
Db  1119 --------TDQEVKAGDKVTF-------KAGKNLKVKQSEKDFTYSLQDTLTGLTSI 1160
Qy  1474 GLSASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGK 1533
Db  1161 TLGGTAN--GRNDTGTVINKDGLTITLAN-------GAAAGTDASNGNTISVTKDGI 1208
Qy  1534 PVGIFPAFTSTNVSAAL-------ALDNRTSQSISLELKRAEPV-----TSNDISELTS 1580
Db  1209 SAG---NKEITNVKSALKTYKDTQNTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTD 1265
Qy  1581 TLGKHFK--DSATTKMLAALKELDDAK-------PAEQLHILQQHFSAKD 1621
Db  1266 NNGKHTVTIDVAEAKVGDGLEKDTDGKIKLKVDNTDGNNLLTVDATKGASVAKGEFNA-- 1323
Qy  1622 VVGDERYEAVRNLK--KLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKH 1678
Db  1324 VTTDATTAQGTNANERGKVVVK---------GSNGATATETDKKKVATVG---------- 1364
Qy  1679 FDAALPASSAKRLGEMMNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGK 1738
Db  1365 -DVAKAINDAATFVKVENDDSATID----DSPTDDGANDALKAXDTL--TLKAGKNLK 1415
Qy  1739 VGREEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSERNIGTINF- 1797
Db  1416 VKRDGKNITFALANDLSVKSATVSDKLS-----------LGT-NGNKVNITSDTKGLNFA 1463
Qy  1798 ---KYGQDQNTPRRFTLEGGIAQANPQVASALTD 1828
Db  1464 KDSKTGDDAN----IHLNG-------IASTLTD 1485
```

```
Qy   296 MLKGSG------------AGVTPL-AVTLDKG--KLQLAPD--NPPALNTLLKQ 332
Db    72 DKEGTGEKEVTENSNWGIYFDNKGVLKAGAITLKAGDNLKXKQXTDEXTNASSFTYSLKK 131
Qy   333 TLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGT 392
Db   132 DLTDLT------SVATEKLSFGANGDKVDITSDANGL----KLAKTGN 169
Qy   393 GSVSVDG------KSGKISLGSGTQSH-NKTMLSQPGEAHRSLLTGIWQHPAGAAR 441
Db   170 GNVHLNGLDSTLPDAVTNTGVLSSSSFTPNDVEKTRAA---TVKDVLNAGW-NIKGAKT 224
Qy   442 PQG--ESIRLHD-DKIHILHPELGVWQSADKDT--------HSQLSR 477
Db   225 AGGNVESVDLVSAYNNVEFI----TGDKNFLDVVLTAKENXKTTEVKFTPKTSVIK 276
Qy   478 QADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSL 537
Db   277 EKDGKLFTGKENNDTNKVTSNFATDN------TD-EGNGLVTAKAVI 316
Qy   538 DASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDG 597
Db   317 DAVNKAGWRVKTTANGQNG------DFATVASGTNVFES-GDGTTASVTKDTNG 365
Qy   598 NELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGW 657
Db   366 NGITVK------------YDAKVGDGLKFD-- 383
Qy   658 NLTDALVIDNQLGLHHTNPEPHEIL-----DMGHLGSLALQEGKLHYFDQLTKGWTG-A 710
Db   384 --SDKKIVADTTALTVGGKVAEIAKEDDKKIVNAGDLVTALGNL--------SWKAKA 433
Qy   711 ESDCKQLKKGLDGAAYLLKDGEV-----KRLNINQS----TSSIKHGTENVFSL 755
Db   434 EADTGALEGIS------KDQEVKAGETVTFKAGKNLKVKQDGANFTYSLQDALTGLTSI 487
Qy   756 PHVRNKPEPGDALQGLNKDDKAQAMA---VIGVNKYLALTEKGDIRSFQIKPGTQQLER 811
Db   488 TLGGTTNGGNDAKTVINKDGLTITPAGNGGTTGTNT-ISVTKDG-----IKAGNKAITN 540
Qy   812 PAQTL------SREGISGELKDI--HYD-------HKQNLYALTHEGEVFHQP 849
Db   541 VASGLRAYDDANFDVLNNSATDLNRHVEDAYKGLLNLNEKNANKQPLVTDSTAATVGDLR 600
Qy   850 REAW------QNGAESSS-----------WHKLALPQSESKLKSLDM 879
Db   601 KLGWVVSTKNGTKEESSNQVKQADEVLFTGAGAATVTSKSENGKHTITVSVAEFTK--ADC 657
Qy   880 SHEHKPIATFEDGSQHLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGS 939
Db   658 GLE------KDGDTIKLKV-----DNQNTDNVLTVGNNGTAVTKG 691
Qy   940 GL-TVKLSAQTGGMTCAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGRE 998
Db   692 GFETVKT-----GATDADRGKVT-----VKDATANDADKKVATVKDVATAINSA------ 735
Qy   999 GLKPLYEMQGALIKQLDAHVRHNAPQPDLQS-----KLETLDLGEHGAELLNDMKRF 1051
Db   736 ---ATFVKTENLTTSIDEDNPTDNGKDDALKAGDTLTFKAGKNLKVKRDGKNITFDLAKN 792
Qy  1052 RDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDL------ 1104
Db   793 LEVKTAKVSDTLTIGGN------TPTGGTTATPKVNITSTADGLNFAKETA 837
Qy  1105 ----SKSLQQ--AVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPL------ 1146
Db   838 DASGSKNVYLKGIATTLTEPSAGAK-----SSHVDLNVDATKKSNAASIEDVLRAGWNIQ 892
Qy  1147 --GRQRD------PNDKTALTFKSRLILDTVTI------GELHELADKAKLVSDHK 1187
Db   893 GNGNNVDYVATYDTVNFTDDSTGTT------TVTVTQKADGKGADVKIGAKTSVTKDHN 945
Qy  1188 PDADQIKLRQQFD--TLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEH 1245
```

```
PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 08/409,995
    FILING DATE:  24-MAR-1995
ATTORNEY/AGENT INFORMATION:
    NAME:  Silva, Robin M.
    REGISTRATION NUMBER:  38,304
    REFERENCE/DOCKET NUMBER:  A-61053-2/RFT/RMS
TELECOMMUNICATION INFORMATION:
    TELEPHONE:  (415) 781-1989
    TELEFAX:  (415) 398-3249
    TELEX:  910 277299
INFORMATION FOR SEQ ID NO: 4:
    SEQUENCE CHARACTERISTICS:
    LENGTH:  1912 amino acids
    TYPE:  amino acid
    STRANDEDNESS:  unknown
    TOPOLOGY:  unknown
    MOLECULE TYPE:  protein
US-08-685-467-4

Query Match      2.2%;    Score 209.5;   DB 3;    Length 1912;
Best Local Similarity  18.9%;   Pred. No. 8.1e-06;
Matches 346;  Conservative 226;  Mismatches 653;  Indels 609;  Gaps 85;
```

```
Qy  236 SEFQQLHQQRLARERENPPQQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQS 295
Db   20 SELIRTHTKRL-RNRGDPV----LATLLFATVQANATDEDEE-LDPVVRTAPVLSFHS    71

Qy  296 MLKGSG----------AGVTPL-AVTLDKG--KLQLAPD--NPPALNTLLKQ           332
Db   72 DKEGTGEKEVTENSNWGIYFDNKGVLKAGAITLKAGDNLKXKQXTDEXTNASSFTYSLKK  131

Qy  333 TLGKDTQHYLAHHASSDGSQHLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGT   392
Db  132 DLTDLT----------SVATEKLSFGANGDKVDITSDANGL------KLAKTGN        169

Qy  393 GSVSVDG-----------KSGKISLGSGTQSH-NKTMLSQPGEAHRSLLTGIWQHPAGAAR 441
Db  170 GNVHLNGLDSTLPDAVTNTGVLSSSSFTPNDVEKTRAA---TVKDVLNAGW-NIKGAKT   224

Qy  442 PQG--ESIRLHD--DKIHILHPELGVWQSADKDT----------HSQLSR            477
Db  225 AGGNVESVDLVSAYNNVEFI----TGDKNTLDVVLTAKENXKTTEVKF'TPKTSVIK     276

Qy  478 QADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSL  537
Db  277 EKDGKLFTGKENNDTNKVTSNTATDN-------TD-EGNGLVTAKAVI              316

Qy  538 DASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDG  597
Db  317 DAVNKAGWRVKTTTANGQNG----------DFATVASCTNVTFES-GDGTTASVTKDING  365

Qy  598 NELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGW  657
Db  366 NGITVK------------------YDAKVGDGLKFD--                        383

Qy  658 NLFDALVIDNQLGLHHTNPEPHEIL-------DMGHLGSLALQEGKLHYFDQLTKGWTG-A 710
Db  384 --SDKKIVADTTALTVTGGKVAEIAKEDDKKKLVNAGDLVTAIGNL------SWKAKA    433

Qy  711 ESDCKQLKKGLDGAAYLLKDGEV-------KRLNINQS---TSSIKHGTENVFSL       755
Db  434 EADTQGALEGIS-----KDQEVKAGETVTFKAGKNLKVKQDGANFTYSLQDALTGLTSI   487

Qy  756 PHVRNKPEPGDALQGLNKDDKAQAMA----VIGVNKYLALTEKGDIRSFQIKPGTQQLER  811
Db  488 TLGGTTNGGNDAKTVINKDGLTITPAGNGGTTGTNT-ISVTKDG-----IKAGNKAITN  540

Qy  812 PAQTL-----SREGISGELKDI--HVD-----------HKQNLYALTHEGEVFHQP      849
Db  541 VASGLRAYDDANFDVLNNSATDLNRHVEDAYKGLLNLNEKNANKQPLVTDSTAATVGDLR  600

Qy  850 REAW-----QNGAESSS------------------------------WHKLALPQSESKLKSLDM 879
Db  601 KLGWVVSTKNGTKEESNQVKQADEVLFTGAGAATVTSKSENGKHTITVSVAETK---ADC      657

Qy  880 SHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGVIPGS   939
Db  658 GLE-----KDGDTIKLKV--------------DNQNTDNVLTVGNNGTAVTKG          691

Qy  940 GL-TVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGRE  998
Db  692 GFETVKT-----GATDADRGKVT----------VKDATANDADKKVATVKDVATAINSA---   735

Qy  999 GLKPLYEMQGALIKQLDAHNVRHNAPQPDLQS------KLETLDLGEHGAELLNDMKRF  1051
Db  736 --ATFVKTENLTTSIDEDNPTDNGKDDALKAGDTLTFKAGKNLKVKRDGKNITFDLAKN   792

Qy 1052 RDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDL----    1104
Db  793 LEVKTAKVSDTLTIGGN-----------------TPTGGTTATPKVNITSTADGLNFAKETA 837

Qy 1105 ---SKSLQQ--AVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPL------------  1146
Db  838 DASGSKNVYLKGIATTLTEPSAGAK---SSHVDLNVDATKKSNAASIEDVLRAGWNIQ   892

Qy 1147 -GRQRD----------PNDKTALTKSRLILDTVTI--------GELHELADKAKLVSDHK  1187
Db  893 GNGNNVDYVATYDTVNFTDDSTGTT---TVTVTQKADGKGADVKIGAKTSVIKDHN      945

Qy 1188 PDADQIKQLRQQFD--TLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEH 1245
Db  946 GKLFTGKDLKDANNGATVSEDDGKD----TGTGLVTAKTV--IDAVNK--SGWRVTG     994

Qy 1246 HGVNLTTRTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPV 1305
Db  995 EGATAETGATAVNAGNAE-------TVTSGTSVNFK---NGNATTATV-             1032

Qy 1306 PVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMP---YMTGKK   1360
Db 1033 -----------SKDNGNINVKY--DVNVGDGLKIGDDKKIVADTTTLTVTGGK         1072

Qy 1361 TSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTIQNSLKFKLTEDELPGFIHGLTHGTLT 1420
Db 1073 VSVPAGANSVNNNKKL----------VNAEGLATALNNLSWTAKADK--YADGESEGE-  1118

Qy 1421 PAELLQKGIEHQMKQGSKLTFSVDTSANLDLRAGINLNEDGSK------PNGVTARVSA  1473
Db 1119 ----------TDQEVKAGDKVTF------KAGKNLKVKQSEKDFTYSLQDTLTGLTSI   1160

Qy 1474 GLSASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGK 1533
Db 1161 TLGGTAN---GRNDTGTVINKDGLTITLAN----------GAAAGTDASNGNTISVTKDGI 1208

Qy 1534 PVGIFPAFTSTNVSAAL----------ALDNRTSQSISLELKRAEPV---TSNDISELTS 1580
Db 1209 SAG----NKEITNVKSALKTYKDTQNTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTD 1265

Qy 1581 TLGKHFK--DSATTKMLAALKELDDAK----------PAEQLHILQQHFSAKD         1621
Db 1266 NNGKHTVTIDVAEAKVGDGLEKDTDGKIKLKVDNTDGNNLLTVDATKGASVAKGEFNA-- 1323

Qy 1622 VVGDERYEAVRNLK--KLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKH  1678
Db 1324 VTTDATTAQGTNANERGKVVVK--------GSNGATATETDKKKVATVG---          1364

Qy 1679 FDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGK 1738
Db 1365 -DVAKAINDAATFVKVENDDSATID---DSPTDDGANDALKAXDTL--TLKAGKNLK    1415

Qy 1739 VGREEVGVLFQDRRNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSERNIGTINF- 1797
Db 1416 VKRDGKNITFALANDLSVKSATVSDKLS----------LGT-NGNKVNITSDTKGLNFA  1463

Qy 1798 --KYGQDNTPRRFTLEGGIAQANPQVASALTD                              1828
```

```
Db   1464 KDSKTGDDAN-----IHLNG-------IASTLTD 1485
            |||  ||         ||  |     :|| ||
```

RESULT 3
US-08-769-309A-5

; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
;   APPLICANT: Scott, John D.,
;   APPLICANT: Nauert, Brian J.,
;   APPLICANT: Klauck, Theresa M.
;   TITLE OF INVENTION: Protein Binding Domains of Gravin
;   NUMBER OF SEQUENCES: 24
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower/233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/769,309A
;   FILING DATE:
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: No. 5741890and, Greta E.
;   REGISTRATION NUMBER: 35,302
;   REFERENCE/DOCKET NUMBER: 27866/33451
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312-474-6300
;   TELEFAX: 312-474-0448
;   TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1780 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-769-309A-5

Query Match   2.2%; Score 203.5; DB 1; Length 1780;
Best Local Similarity 17.8%; Pred. No. 1.9e-05;
Matches 338; Conservative 279; Mismatches 717; Indels 565; Gaps 81;
```

```
Qy    29 QQGSSSSSPQNAAASLAAEGKNRGKMPRIHQPSTAADG-----ISAAHQQKKSFSLRGCL  83
            :   :  ::  :  :   :  ||  :    :    :::
Db   245 EQSHAEISPPAESGQAVEECKEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQGWA 304

Qy    84 GTKK---FSRSAPQQGPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLTRSGGVKRRN 140
          ||::     :  :  ||::  |  ||||   | |:      : ||   | :|:::|
Db   305 GWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGK-----AEVASEKLTASEQAHPQE 359

Qy   141 MDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPH 200
            : ::       ::       |   :  ||    | :   :|
Db   360 PAESAHEPRLSAEYEKVELPSEEQ------------------------------------ 383

Qy   201 IPGSHHEIKEPVGSTSKATTAHADRVEIAQEDDSEF------QQLHQQRLARERE--N 252
          : :::: :         |    |           :::
Db   384 VSGSGQPSEEKP--APLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQKTEVEETAGS 440

Qy   253 PPQPPKLGV-ATPISARFQPKLTAVAESVLEGTDTTQ-SPLKP-QSMLKGSSGAGV---TP 306
          : |::   |  :::| | :|       ::: :|:   ::|:|     |:|        ||
Db   441 VPAEELVGMDAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVE 500

Qy   307 LAVTLDKGKLQLAPDNPPALNTLLKQTLGK-------------DTQHYLAHHASSDGSQHLL 355
          :   :::  : :|  | ::  :      ||             :: :|   ||| ||
```

```
Db   501 MLSSQERMKVQGSPLKKLFTSTGLKKLSGKKKQKGKRGGDEESGEHTQVPADSPDSQE-- 558
           ||| |   ||                      |||  ||
Db   356 LDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNK 415
          :||: ::       :||:  :   :
Qy   559 -EQKG---ESSASS-----PEEPEEI--TCLEKGLAEVQQDGEAEE--GATSDGEKK 602
                :  :: || :|
Qy   416 TMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQL 475
          :|:  :        :                       :|||
Db   603 REGVTPWASFKKMVT-----PKKRVRRPSES------DKEDELDKVKSATLSSTESTASEM 652
          |:|       ::  :::                      |
Qy   476 SRQADGKLYALKDNRTLQNLSDNKSSEKLV---------DKIKSYSVDQRGQVAI-----L 522
          :  ||  : |: :|                             :|::
Db   653 QEEMKGSVEEPKPEPKRKVDTSVSWEALICVGSSKKRARRSSSDEEGGPKAMGGDHQK 712
          ||| ||             |:::    :|         |||  ||
Qy   523 TDTPGRHKMSIMPSLDASPESHISLSLHFADAHQG-----LLHGKSELEAQSVAISHGRL 577
          ||:||  |:                                ::|:|
Db   713 ADEAGKDKETGTDGILAGSQEH------DPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764
          ||    |  ||                   |   |: || || | | ||
Qy   578 VVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN 637
          ::  ||   : |               |  :||::     :
Db   765 VT---------PRKKSKSKLEEKS-----EDSIAG------- 785
          ||          ||::| ||
Qy   638 NFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNEPEHEILDMGHLGSLALQEGKL 697
            |  |  |            |    |:  :    |:||
Db   786 ----------------------------SGVEHSTPDTE--------- 796
                                      :|  |:|:|:
Qy   698 HYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPH 757
          |: :|: |               ||       |:| :   :  |
Db   797 ----------PGKEESWYSIKKFIPGRRKKRPDGKQEQAPVEDAGPT---------- 833
                     | |||: :||||  |||||||  |
Qy   758 VRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTLS 817
          :|||||   | :: |          ||:|:|::|:  |  :|||| ||  :  |
Db   834 ----------GANEDD-SDVPAVVPLSEYDAV-EREKMEAQAQKGAEQPEQKAAT- 877
                      || ||     |:||:|     |:||    |    |   | |
Qy   818 REGISGELKDIHVDHKQNLYA-LTHEGEVFHQPREAWQNGAESSSWHKL--ALPQSESK 873
          :|     |      | ::          |     :    ||:  :||:    |:|:||
Db   878 --EVSKELSESQVHMMAAAVADGTRAATIEERSPSWISASVTEPLEQVEAEAALLTEEV 935
             :|  | |  |:||       |  |  |::||  |:   :|: || :|| |  ||
Qy   874 LKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKG 933
          ||:  |::||:|  | :||:     || :|           ||:  |||: :  |||  |
Db   936 LEREVIAEEEPPIVT-----------NREARG--D 962
          ||||:|||||: |                |||:|  |
Qy   934 KVIPGSGLTVK--LSAQTGGMTGA-EGRRVSSKFSERIRAYAFNPTMSTPRPIKNAAYAT 990
          |:|||: | |    ||  | ||||  ||::|||||:|||||||:|::  | |||||||||
Db   963 TVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSAVSQLTDSPDTTEAT---- 1019
          |::|| ||| || |||| ||| ||  || ||| |||: || | |||:| ||:|||
Qy   991 QHGWQGREGLKPLYEMOGAL-------------IKQLDAHNVRHNAPQPDLQSKLETLDLG 1038
          |||||| | |  ::| |              :|:| |:    ||  ||| ||:||:|| |
Db  1020 ----------------------PVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLP--G 1057
                                ||||| |||:|||||||||||||||||||||:|||||| |
Qy  1039 EHGAE-LLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSP--GKALVQS 1094
          ||||:  |||||||||||:|||| ||||| ||| ||| | :||:|||||   ||||| |
Db  1058 TGGPEDVLQPVQRAEAERPEEQAEASGLKKETDVVLKVDAQ--EAKTEPFTQGKVV--- 1111
          ||| || ||:| |:|| ||||| || |:| |:|||:|||   |||:|| ||| |
Qy  1095 FNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPND 1154
          |:|||| | |:|:| | |  ||:|||:|:|:| |||||||||| |:|| ||| |:| ||
Db  1112 ------GQTTPESFEKAPQVTESIESSELVTTCQAETLAGV-----KSQEMVMEQAIPPD 1160
                | || ||:|| || |||:||||| | |||| ||||     || || ||| |||
Qy  1155 KTALFKSRLLIDLTVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVK 1214
          ||| ||| |::|  ||::|  |  || ||  ||| |||:| ||| |||| | |::| |:|
Db  1161 S-------VETPTDSETDGSTPVADFDAPGTTQKDEIVEI-------HEENEV-------- 1199
                  |||  ||:|  | |  |||  ||  |:|| |        ||||||
Qy  1215 HYTDMGFTHNKALEANYD--AVKAFINAFKKEHHGVNLTTRTVLESOGSAELAKKLKNT 1271
          ||  |  |:|| || |    |:|  || |:||:| ||| || ||:|| |||  | |:|
Db  1200 HLVPVRGTEAEAVPAQKERPPAPSSFV--FQEE------------TKEQSKMEDT 1240
          || |:| |||  |   ||  | || |   ||||            |:|| |:||
Qy  1272 LLSLD---SGESMS-FSRSYGGGVSTVFVPTLSKKVPVPFVIPGAGITLDRAYNLSFSRTS 1327
          |||     |||||  ||| |||||||| |||||||||| || ||||| |||||| || |
Db  1241 LEHTDKEVSVETVSIISKTEGTQEADYADEKTKDVPF--------FE------------1280
          ||| |:|||| || :|| || ||| |    |:| ||           |
Qy  1328 GGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNA--SDWLSAK-HKISPDLRIGA 1384
          || ||| | |||| |:|  | | | ||||| |::| |     |  |   | |||| | |
Db  1281 -----------GLEGSIDTGITVSREKVTEVALKGEGTEEAECKKDDALELQSHAKSP-----1327
                     || || | :| || :||    |  | || |   || |  |  | |:|
```

```
; MOLECULE TYPE: protein
US-08-994-570-5

Query Match          2.2%;  Score 203.5;  DB 3;  Length 1780;
Best Local Similarity   17.8%;  Pred. No. 1.9e-05;
Matches  338;  Conservative  279;  Mismatches  717;  Indels  565;  Gaps  81;

Qy    29  QQGSSSSSPQNAAASLAAEGKNRGKMPRIHQPSTAADG----ISAAHQQKKSFSLRGCL   83
                :   :   :   ::  :  ::    :     :  :||||
Db   245  EQSHAEISPPAESGQAVEECKEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQGWA  304

Qy    84  GTKK--FSRSAPQGQPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLTRSGGVKRRN  140
              :    ::  :  : ||  :  :  :  ::   |
Db   305  GWRKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGK---AEVASEKLTASEQAHPQE  359

Qy   141  MDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPH  200
                                                                    :: :
Db   360  PAESAHEPRLSAEYEKVELPSEEQ------------------------------------  383

Qy   201  IPGSHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEF----QQLHQQRLARERE--N  252
          : ||   :     :  : : ::   :  :  :  :
Db   384  VSGSQGPSEEKP---APLATEVFDEKIEVHQEEVVAEVHVSTVEERTEQKTEVEETAGS  440

Qy   253  PPQPPKLGV-ATPISARFQPKLTAVAESVLEGTDTQ-SPLKP-QSMLKGSGAGV---TP  306
            :   : :     :  :     :  :   :
Db   441  VPAEELVGMDAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVSEVE  500

Qy   307  LAVTLDKGKLQLAPDNPPALNTLLKQTLGK-----------DTQHYLAHHASSDGSQHLL  355
           : :   :  :   :||: ||  : ||
Db   501  MLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKRGGGDEESGEHTQVPADSPDSQE--  558

Qy   356  LDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNK  415
               :
Db   559  -EQKG-----ESSASS----PEEPEEI-TCLEKGLAEVQQDGEAEE---GATSDGEKK  602

Qy   416  TMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQL  475
           :  :     :  :  : |  :   : : ||
Db   603  REGVTPWASFKKMVT----PKKRVRRPSES----DKEDELDKVKSATLSSTESTASEM  652

Qy   476  SRQADGKLYALKDNRTLQNLSDNKSSEKLV-----DKIKSYSVDQRGQVAI----L  522
            :  :  : :     : :  |     :  : ||
Db   653  QEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRRSSDEEGGPKAMGGDHQK  712

Qy   523  TDTPGRHKMSIMPSLDASPESHISLSLHFADAHQG---LLHGKSELEAQSVAISHGRL  577
              :  : |  :     : :  : ||
Db   713  ADEAGKDKETGTDGILAGSQEH------DPGQGSSSPEQAGSPTEGEGVSTWESFKRL  764

Qy   578  VVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN  637
           :   : :  :  : |   :     : : :  |
Db   765  VT--------PRKKSKSKLEEKS--------EDSIAG------  785

Qy   638  NFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHELDMGHLGSLALQEGKL  697
                                              :
Db   786  --------------------------------SGVEHSTPDTE----  796

Qy   698  HYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPH  757
               : |    :                 | |
Db   797  --------PGKEESWVSIKKFIPGRRKKRPDGKQEQAPVEDAGPT-----  833

Qy   758  VRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLS  817
           :   |  : :  :  ::: :  : :::   |  :
Db   834  ----------GANEDD-SDVPAVVPLSEYDAV-EREKMEAQQAQKGAEQPEQKAAT---  877

Qy   818  REGISGELKDIHVDHKQNLYA-LTHEGEVFHQPREAWQNGAESSSWHKL---ALPQSESK  873
           :  : : :: |  :   :  | :    :      :    :::
Db   878  --EVSKELSESQVHMMAAAVADGTRAATIEERSPSWISASVTEPLEQVEAEAALLTEEV  935

Qy   874  LKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKG  933
           :  :    : :  :  :   :  :  |::
Db   936  LEREVIAEEEPPTVT----------------NREARG--D  962

Qy   934  KVIPGSGLTVK--LSAQTGGMTGA-EGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYAT  990
```

```
Qy  1385  AVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVD  1444
              ::   :   :|| :    ||   :  :   :|| :|
Db  1328  --PSPVEREMVVQVEREKTEAE----------------PTHVNEEKLEHE---TAVTVSEE  1367

Qy  1445  TSANLDLRAGINLNE--------DGSKP----NGVTARVSAGLS-ASANLAAGSRERS  1489
           :   :   ::   |||:  ::  :  :||  :   :| |:
Db  1368  VSKQLLQTVNPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEK  1427

Qy  1490  TTSGQFGSTTSASNRPTFLNGVGAGANLTAALGVAHS----STHEGK---PVG-----  1536
           || ::  :|||:|    :|||:  :  :  :     ::
Db  1428  V----LGETANILETGETL--EPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAK  1480

Qy  1537  ----IFPAFTSTNVSAALALDNRTS-------------QSISLELKRAEPVTSNDIS  1576
               :  : |:   :|| :  ::   :|| : :   :: ::   |
Db  1481  STPVTVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEVKVSVAIEDLEP--ENGIL  1538

Qy  1577  ELTSTLGKHFK-----------DSATTKMLAALKELDDAKPAEQLHIL---QQHFSA  1619
           || :  :| :||           :| :    :  :  :|:   |
Db  1539  ELETKSSKLVQNIIQTAVDQFVRTETATEMLTS-ELQ----TQAHVIKADSQDAGQE  1591

Qy  1620  KDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHF  1679
           :  :  :  ::   :     :: :   :   :  :|     :  ::
Db  1592  TEKEGEEPQASAQD--ETPITSAKEESESTAVGQA-HSDISKDMS------EASEKTM  1640

Qy  1680  DAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDG--LREQTEKAILD  1736
             |:     :| ::   :: :       :      ::  |   :|:  |: : : :||
Db  1641  TVEVPGSTV-------NDQQLEEVVLPSEEEGGAGTKSVPEDDGHALLAERIEKSLVE  1692

Qy  1737  GKVGREEVGVLFQDRNNLRVKSVSVSQSVSK-SEGFNTP  1774
           |:    :  : |   :: :  :     :  : || :|||
Db  1693  PKEDEKGDVDDPENQNSALADTDASGGLTKESPDTNGP  1731
```

```
```

```
;   FILE REFERENCE:  065064/0128
;   CURRENT APPLICATION NUMBER:  US/09/377,155
;   CURRENT FILING DATE:  1999-08-19
;   PRIOR APPLICATION NUMBER:  PCT/AU98/01031
;   PRIOR FILING DATE:  1998-12-14
;   PRIOR APPLICATION NUMBER:  GB 9726398.2
;   PRIOR FILING DATE:  1997-12-12
;   NUMBER OF SEQ ID NOS:  33
;   SOFTWARE:  PatentIn Ver. 2.0
;   SEQ ID NO 33
;   LENGTH:  2353
;   TYPE:  PRT
;   ORGANISM:  Haemophilus influenzae
US-09-377-155-33

Query Match                    2.1%;   Score 201;  DB 4;   Length 2353;
Best Local Similarity   18.9%;  Pred. No. 4.5e-05;
Matches 346;   Conservative 223;   Mismatches 223;   Indels 665;   Gaps 592;   84:

Qy   236  SEFQOLHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQS  295
               |: :|  | :||  ::|    :: |  | ::|   |:|
Db    20  SELTRTHTKRASATVETAV------LATLLFATVQANATDEDE-LDPVVRTAPVLSFHS   72

Qy   296  MLKGSG-----------------------AGVTPL-AVTLDKGKIQLAPDNPPALNTLLKQTLGKDT  338
               :||                     ::|
Db    73  DKEGTGEKEVTENSNWGIYFDNKGVLKAGAITLKAGD------------NLKIKQNTDEST  121

Qy   339  QHYLAHHASSDGSQHLLLDNKGHLFDIKSTAT-SYSVLHNSHPGEIKG---KLAQAGT    392
               :||                           | |||  ::|       ||
Db   122  -----NASS-----FTYSLKKDLTDLTSVATEKLSFGANGDKVDITSDANGLKLAKTGN   170

Qy   393  GSVSVDG------------KSGKISLGSGTQSH-NKTMLSQPGEAHRSLLTGIWQHPAGAAR   441
               |:|:|               ::|:  |:               | |:  ||
Db   171  GNVHLNGLDSTLPDAVTNTGVLSSSSFTPNDVEKTRAA----TVKDVLNAGW-NIKGAKT   225

Qy   442  PQG--ESIRL----------IHILHPELGVWQSADKDTHSQLSRQADGKLYA   485
               |:|  :|            |||:  |        : |: | |:|:::  ||:|
Db   226  AGGNVESVDLVSAYNNVEFITGDKNTLDVVLTAKENGKTTEVKFTPKTSVIKEKDGKLFT   285

Qy   486  LKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILDTPGRHKMSIMPSLDASPESHI   545
               |                     :|  | :::|::  |     |:
Db   286  GKENNDTNKVTSNTATDN------------TD-EGNGLVTAKAVIDAVNKAGW        325

Qy   546  SLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAM   605
               ||     ||  ||| |:    ::                 |
Db   326  RVKTTANGQNG--------DFATVASGTNVTFES-GDGTTASVTKDTNGNGITVK-----  372

Qy   606  PQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVI   665
Db   373  --------------------------------------------------YDAKVDGLKFD-----  390

Qy   666  DNQLGLHHTNPEPHEIL-----DMGHLGSLALQEGKLHYFDQLTKGWTG-AESDCKQLK   718
                 ||                   ::   ||
Db   391  ADTTALTVTGGKVAEIAKEDDKKKLVNAGDLVTALGNL----SWKAKAEADTDGAL     442

Qy   719  KGLDGAAYLLKDGEV-----------KRLNINQS---VIGVNKYLALTHEGEVFHQPREAW---  763
Db   443  EGIS-----KDQEVKAGETVTFKAGKNLKVKQDGANFTYSLQDALTGLTSITLGGTTNG      496

Qy   764  PGDALQGLNKDDKAQAMA-----VIGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTL--      816
                ||:  || ::|                :| |: ||
Db   497  GNDAKTVINKDGLTITPAGNGGTTGTNT-ISVTKDG------IKAGNKAITNVASGLRAY   549

Qy   817  --SREGISGELKDI--HVD---------------------HKQNLYALTHEGEVFHQPREAW--   853
Db   550  DDANFDVLNNSATDLNRHVEDAYKGLLNLNEKNANKQPLVTDSTAATVGDLRKLGWVVST      609

Qy   854  QNGAESSS---------WHKIALPQSESKLKSLDMSHEHKPIA   887
                ||| ::|         :||:
Db   610  KNGTKEESNQVKQADEVLFTGAGAATVTSKSENGKHTITVSVAETK---ADCGLE---      661

Qy   888  TFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGL-TVKLS   946
```

```
Db   963  TVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSAVSQLTDSPDTTEAT---  1019
Qy   991  QHGWQGREGLKPLYEMQGAL----------IKQLDAHVRHNAPQPDLQSKLETLDLG     1038
Db  1020  --------PVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLP---------G       1057
Qy  1039  EHGAE-LLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSP--GKALVQS   1094
Db  1058  TGGPEDVLQPVQRAEAERPEEQAEASGLKKETDVVLKVDAQ--EAKTEPFTQGKVV---  1111
Qy  1095  FNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPND  1154
Db  1112  --------GQTTPESFEKAPQVTESIESSELVTTCQAETLAGV---KSQEMVMEQAIPPD  1160
Qy  1155  KTALTKSRLLLDTVTIGELHELADAKKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVK  1214
Db  1161  S----------VETPTDSETDGSTPVADFDAPGTTQKDEIVEI---------HEENEV-  1199
Qy  1215  HYTDMGFTHNKALEANYD--AVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNT   1271
Db  1200  HLVPVRGTEAEAVPAQKERPAPSSFV-FQEE-----------TKEQSKMEDT         1240
Qy  1272  LLSLD--SGESMS-FSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYNLSFSRTS   1327
Db  1241  LEHTDKEVSVETVSILSKTEGTQEADQYADEKTKDVPF------FE----          1280
Qy  1328  GGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNA--SDWLSAK-HKISPDLRIGA  1384
Db  1281  -----GLEGSIDTGITVSREKVTEVALKGEGTEEAECKKDDALELQSHAKSP-------  1327
Qy  1385  AVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVD  1444
Db  1328  --PSPVEREMVVQVEREKTEAE-----------PTHVNEEKLEHE---TAVTVSEE     1367
Qy  1445  TSANLDRAGINLNE-----------DGSKP------NGVTARVSAGLS-ASANLAAGSRERS  1489
Db  1368  VSKQLLQTVNPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTRAAAEEEK  1427
Qy  1490  TTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHS---STHEGK----PVG-----  1536
Db  1428  V----LGETANILETGETL---EPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAK  1480
Qy  1537  ----IFPAFTSTNVSAALALDNRTS----------QSISLELKRAEPVTSNDIS        1576
Db  1481  STPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEVKVSVAIEDLEP--ENGIL  1538
Qy  1577  ELTSTLGKHFK-----------DSATTKMLAALKELDDAKPAEQLHIL---QQHFSA   1619
Db  1539  ELETKSSKLVQNIIQTAVDQFVRTEETATEMLTS--ELQ----TQAHVIKADSQDAGQE  1591
Qy  1620  KDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHF  1679
Db  1592  TEKEGEEPQASAQD--ETPITSAKEESESTAVGQA-HSDISKDMS------EASEKTM    1640
Qy  1680  DAALPASSAKRLGEMNNDPALKDIIKQLQSTPFSSASVSMELKDG---LREQTEKAILD   1736
Db  1641  TVEVEGSTV----------NDQQLEEVLPSEEEGGGAGTKSVPEDDGHALLAERIEKSLVE  1692
Qy  1737  GKVGREEVGVLFQDRNNLRVKSVSVSQSVSK-SEGFNTP  1774
Db  1693  PKEDEKGDDVDDPENQNSALADTDASGGLTKESPDTNGP  1731
```

RESULT 5
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN

US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
;   STREET: Four Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: United States
;   ZIP: 94111-4187
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/913,942
;   FILING DATE: 29-DEC-1997
;   CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/409,995
;   FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US96/4031
;   FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
;   NAME: Vance, Dolly A.
;   REGISTRATION NUMBER: 39,054
;   REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 781-1989
;   TELEFAX: (415) 398-3249
;   TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2353 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-913-942-4

Query Match   2.1%;   Score 201; DB 4; Length 2353;
Best Local Similarity 18.9%; Pred. No. 4.5e-05;
Matches 346; Conservative 223; Mismatches 665; Indels 592; Gaps 84;

Qy   236 SEFQQLHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQS 295
                ||·|| | ||·||||||||||||||| || ||||||||·||·||·:
Db   20 SELTRTHTKRASATVETAV---------LATLLFATVQANATDEDEE-LDPVVRTAPVLSFHS   72

Qy   296 MLKGSG------------------AGVTPL-AVTLDKGKLQLAPDNPPALNTLLKQTLGKDT 338
                ·||:| |·||| || :|| ||
Db   73 DKEGTGEKEVTENSNWGIYFDNKGVLKAGAITLKAGD----------NLKIKQNTDEST 121

Qy   339 QHYLAHHASSDGSQHLLLDNKGHLFDIKSTAT-SYSVLHNSHPGEIKG-----KLAQAGT 392
                : | | · | |||·| ||||·||||·||·
Db   122 -------NASS----FTYSLKKDLTDLTSVATEKLSFGANGDKVDITSDANGLKLAKTGN 170

Qy   393 GSVSVDG----------------KSGKISLSGSGTQSH-NKTMLSQPGEAHRSLLTGIWQHPAGAAR 441
           |·||·||:| ||||·||||||||||
Db   171 GNVHLNGLDSTLPDAVTNTGVLSSSSFTPNDVEKTRAA----TVKDVLNAGW-NIKGAKT 225

Qy   442 PQG--ESIRL-------HDDK----IHILHPELGVWQSADKDTHSQLSRQADGKLYA 485
                || ·   · |||| ::|||·|:||||||||||||||||||||||||
Db   226 AGGNVESVDLVSAYNNVEFITGDKNTLDVVLTAKENGKTEVKFTPKTSVIKEKDGKLFT 285

Qy   486 LKDNRTLQNLSDNKSSEKLVDKIKSYSVSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHI 545
                ·|| | | · ||| || || || ::

Db   662 :|| :||
                --KDGDTIKLKV-----------------DNQNTDNVLTVGNNGTAVTKGGFETVKT- 699

Qy   947 AQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEM 1006
                | |: ||:·|| ||·||
Db   700 -GATDADRGKVT---VKDATANDADKKVATVKDVATAINSA-------ATFVK 741

Qy   1007 QGALIKQLDAHNVRHNAPQPDLQS------KLETLDLGEHGAELLNDMKRFRDELEQSA 1059
                ||·| | | || ||:||
Db   742 TENLTSIDEDNPTDNGKDDALKAGDTLTFKAGKNLKVKRDGKNITFDLAKNLEVKTAKV 801

Qy   1060 TRSVTVLGQHQCVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDL----------SKSL 1108
                |:: | | ·| |· ||:||
Db   802 SDTLTIGGN------TPTGGTTATPKVNITSTADGLNFAKETADASGSKNV 846

Qy   1109 QQ--AVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPL------GRQRD- 1151
                || | ||·||: ||
Db   847 YLKGIATTLTEPSAGAK---SSHVDLNVDATKKSNAASIEDVLRAGWNIQGNNNVDY 901

Qy   1152 ------PNDKTALTKSRLILDTVTI-----GELHELADKAKLVSDHKPDADQIKQ 1195
                      ||| | | | || ||
Db   902 VATYDTVNFTDDSTGTT------TVTVTQKADGKGADVKIGAKTSVIKDHNGKLFTGKD 954

Qy   1196 LRQQFD--TLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTR 1253
                |:: | · · ||| | |·| |||
Db   955 LKDANNGATVSEDDGKD------TGTGLVTAKTV---IDAVNK--SGWRVTGEGATAETG 1003

Qy   1254 TVLESQGSAELAKKLKKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGI 1313
                :·||| :|||||
Db   1004 ATAVNAGNAE----------TVTSGTSVNFK---NGNATTATV----------- 1033

Qy   1314 TLDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMP-----YMTGKKTTSAGNASD 1368
                || | ·|| | |· ||
Db   1034 ------------SKDNGNINVKY--DVNVGDGLKIGDDKKIVADTTTLTVTGGKVSVPAGAN 1081

Qy   1369 WLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKG 1428
                · ||·:   ·:| || |||||
Db   1082 SVNNNKKL----VNAEGLATALNNLSWTAKADK--YADGESEGE----------- 1119

Qy   1429 IEHQMKQGSKLTFSVDTSANLDLRAGINLNEDGSK------PNGVTARVSAGLSASANL 1481
                :·:|||| |||||·:| ||||||·|| ·|·
Db   1120 TDQEVKAGDKVTF------KAGKNLKVKQSEKDFTYSLQDTLTGLTSITLGGTAN- 1168

Qy   1482 AAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIPPAF 1541
                | | ||| |:   ·| |·|| ||·|
Db   1169 --GRNDTGTVINKDGLTITLAN----GAAAGTDASNGNTISVTKDGISAG---NK 1214

Qy   1542 TSTNVSAAL-------ALDNRTSQSISLELKRAEPV--TSNDISELTSTLGKHFK- 1587
                |||·|·| | |||| ||:||||||·|||·||||| ||·|
Db   1215 EITNVKSALKTYKDTQNTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTDNNGKHTVT 1274

Qy   1588 -DSATTKMLAALKELDDAK-----------PAEQLHILQQHFSAKDVVGDERYE 1629
                 |·||· |||| | || ||·|| |·||
Db   1275 IDVAEAKVGDGLEKDTDGKIKILKVDNTDGNNLLTVDATKGASVAKGEFNA--VTTDATTA 1332

Qy   1630 AVRNLK---KLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPAS 1686
                ·|:||: ||·|·| |·||| |||·| ||||||
Db   1333 QGTNANERGKVVVK-------GSNGATATETDKKKVATVG-------DVAKAIN 1372

Qy   1687 SAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLRECTEKAILDGKVGREEVGV 1746
                || :| ||||| || ||·|:||
Db   1373 DAATFVKVENDDSATID----DSPTDDGANDALKAGDTL---TLKAGKNLKVKRDGKNI 1424

Qy   1747 LFQDRRNLRLVKSVSVSQSVSKSEGFNTPALLGTSNSAAMSMERNIGTINF---KYCQD 1802
                |· ||·||| ·: || || | ||·|||
Db   1425 TFALANDLSVKSATVSDKLS------LGT-NGNKVNITSDTKGLNFAKDSKTGDD 1472

Qy   1803 QNTPRRFTLEGGIAQANPQVASALTD 1828
                · |·:|||
Db   1473 AN----IHLNG------IASTLTD 1486

RESULT 6

```
Db   286 GKENNDTNKVTSNTATDN-----------------------TD-EGNGLVTAKAVIDAVNKAGW 325
Qy   546 SLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAM 605
Db   326 RVKTTTANGQNG----------DFATVASGTNVTFES-GDGTTASVTKDTNGNGITVK- 372
Qy   606 PQHALDEHFGHDHQISGFFHDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVI 665
Db   373 --------------------------YDAKVGDGLKFD-----SDKKIV 390
Qy   666 DNQLGLHHTNPEPHEIL------DMGHLGSLALQEGKLHYFDQLTKGWTG-AESDCKQLK 718
Db   391 ADTTALTVTGGKVAEIAKEEDDKKKLVNAGDLVTALGNL--------SWKAKAEADTDGAL 442
Qy   719 KGLDGAAYLLKDGEV----------KRLNINQS-----TSSIKHGTENVFSLPHVRNKPE 763
Db   443 EGIS------KDQEVKAGETVTFKAGKNLKVKQDGANFTYSLQDALTGLTSITLGGTTNG 496
Qy   764 PGDALQGLNKDDKAQAMA----VIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTL--- 816
Db   497 GNDAKTVINKDGLTITPAGNGGTTGTNT-ISVTKDG------IKAGNKAITNVASGLRAY 549
Qy   817 ---SREGISGELKDI--HVD-----------HKQNLYALTHEGEVFHQPREAW------ 853
Db   550 DDANFDVLNNSATDLNRHVEDAYKGLLNLNEKNANKQPLVTDSTAATVGDLRKLGWVVST 609
Qy   854 QNGAESSS----------------WHKLALPQSESKLKSLDMSHEHKPIA 887
Db   610 KNGTKEESNQVKQADEVLFTGAGAEAATVTSKSENGKHTITVSVAETK---ADCGLE--- 661
Qy   888 TFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKVIPGSGL-TVKLS 946
Db   662 --KDGDTIKLKV----------------DNQNTDNVLTVGNNGTAVTKGGFETVKT- 699
Qy   947 AQTGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEM 1006
Db   700 ----GATDADRGKVT---------VKDATANDADKKVATVKDVATAINSA------ATFVK 741
Qy  1007 QGALIKQLDAHNVRHNAPQPDLQS------KLETLDLGEHGAELLNDMKRFRDELEQSA 1059
Db   742 TENLLTSIDEDNPTDNGKDDALKAGDTLTFKAGKNLKVKRDGKNITFDLAKNLEVKTAKV 801
Qy  1060 TRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDL-----------SKSL 1108
Db   802 SDTLTIGGN-----------TPTGGTTATPKVNIITSTADGLNFAKETADASGSKNV 846
Qy  1109 QQ---AVHATPPSAESKIQSMLGHFVSAGVDMSHQKGEIPL---------GRQRD- 1151
Db   847 YLKGIATTLTEPSAGAK----SSHVDLNVDATKKSNAASIEDVLRAGWNIQGNGNNVDY 901
Qy  1152 ---------PNDKTALTKSRLILDTVTI------GELHELADKAKLVSDHKPDADQIKQ 1195
Db   902 VATYDTVNFTDDSTGTT------TVTVTQKADGKGADVKIGAKTSVIKDHNGKLFTGKD 954
Qy  1196 LRQQFD--TLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTR 1253
Db   955 LKDANNGATVSEDDGKD-----TGTGLVTAKTV--IDAVNK-SGWRVTGEGATAETG 1003
Qy  1254 TVLESQGSAELAKKLKNTLLSLSDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGI 1313
Db  1004 ATAVNAGNAE---------TVTSGTSVNFK-NGNATTATV----------- 1033
Qy  1314 TLDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMP-----YMTGKKTSAGNASD 1368
Db  1034 --------SKDNGNINVKY-DVNVGDGLKIGDDKKIVADTTTLTVTGGKVSVPAGAN 1081
Qy  1369 WLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKG 1428
Db  1082 SVNNNKKL-------VNAEGLATALNNLSWTAKADK---YADGESEGE----------- 1119
Qy  1429 IEHQMKQGSKLTFSVDTSANLDRAGINLNEDGSK-------PNGVTARVSAGLSASANL 1481
Db  1120 TDQEVKAGDKVTF-----------KAGKNLKVKQSEKDFTYSLQDTLTGLTSITLGGTAN- 1168

Qy  1482 AAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAF 1541
Db  1169 --GRNDTGTVINKDGLTITLAN-----------GAAAGTDASNGNTISVTKDGISAG---NK 1214
Qy  1542 TSTNVSAAL------ALDNRTSQSISLELKRAEPV-----TSNDISELTSTLGKHFK- 1587
Db  1215 EITNVKSALKTYKDTQNTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTDNNGKHTVT 1274
Qy  1588 -DSATTKMLAALKELDDAK--------------PAEQLHIIQQHFSAKDVVGDERYE 1629
Db  1275 IDVAEAKVGDGLEKDTDGKIKLKVDNTDGNNLLTVDATKGASVAKGEFNA--VTTDATTA 1332
Qy  1630 AVRNLK---KLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPAS 1686
Db  1333 QGTNANERGKVVVK---------GSNGATATETDKKKVATVG--------DVAKAIN 1372
Qy  1687 SAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVGREEVGV 1746
Db  1373 DAATFVKVENDDSATID----DSPTDDGANDALKAGDTL--TLKAGKNLKVKRDGKNI 1424
Qy  1747 LFQDRNNLRVKSVSVSQSVSKSEGFNTPALLGTSNSAAMSERNIGTINF---KYGQD 1802
Db  1425 TFALANDLSVKSATVSDKLS-----------LGT-NGNKVNITSDTKGLNFAKDSKTGDD 1472
Qy  1803 QNTPRRFTLEGGIAQANPQVASALTD 1828
Db  1473 AN----IHLNG------IASTLTD 1486
```

```
Query Match          2.0%;  Score 190.5;  DB 1;  Length 1536;
Best Local Similarity  18.6%;  Pred. No. 0.00013;
Matches  297;  Conservative  221;  Mismatches  559;  Indels  521;  Gaps  72;

Qy  304  VTPLAVTLDK----------GKLQLAPDNPPALNTLLKQTLGKDTQHYLAH--HASSD  349
              |  |  :||           ||  |  ||  :||
Db  149  INPNGITIGKDAIINGFTASTLDISNENIKARNFTFEQTKDKALAEIVNHGLITVGKD  208

Qy  350  GSQHLL------LDNKG-----------HLFDIKSTATSYSVLHNSHPGEIKGKLAQ  389
         ||  ||      :||            ||:||
Db  209  GSVNLIGGKVKNEGVISVNGGSISLLAGQKITISDIINPTITYSIAAPENEAVNLGDIFA  268

Qy  390  AGTGSVSVDG---KSGKISLGSGT-QSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQ  443
          |:||         |||  :|  ||   :||  |||
Db  269  KG-GNINVRAATIRNQGKLSADSVSKDKSGNIVLSAKEGEA------------------  308

Qy  444  GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQ----NLSDN  498
                                                 ||| ||       || :|
Db  309  -------------------EIGGVISA------QNQQAKGGKLMTGDKVTLKTGAVIDLSGK  346

Qy  499  KSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGL  558
         ||:| :    | :|            |:|:|
Db  347  EGGETYL------GGDERGE------GKNGIQLAKKTSLEKGSTINVS------------  382

Qy  559  LHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDH  618
            ||    ||:|||:|            :|||
Db  383  --GK----EKGGRAIVWGDIALIDGN------INAQGSGDIAKTGGFV----ETSGHD-  424

Qy  619  QISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT-DALVIDNQLGLHHTNPE  677
                 | :      |  ||        : : |   | :   | : |
Db  425  ---LPIKDNAIVDA-------------KEWLLDFDNVSINAETAGRSNTSE  459

Qy  678  PHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLN  737
                :  |||    :||  ||  |     |||  :|  |||
Db  460  DDEYTGSGNSASTPKRNK----EKTLTNTTLES------------ILKKGTFVNIT  500

Qy  738  INQ-----STSSIKHGTENVFSL----PHVRNKPEPGDALQGLNKDDKAQAMAVIGVN  786
         :|      |  :| :||            ||    ||  |
Db  501  ANQRIYVNSSINLSNGSLTLWSEGRSGGGVEINDITTGDDTRGANL--TIYSGGWVDVH  558

Qy  787  KYLALTEKGDI------RSFQIKPGTQQLERPAQTLSREGISGELKDIHVDH----------  832
         |  : |: :|:       :    ||        :|  ||  ||
Db  559  KNISLGAQGNINITAKQDIAFEKGSNQVITGQGTIT----SGNQKGFRFNNVSLNGTGSG  614

Qy  833  -----KQNLYALTHEGEVFHQPREAWQNGAESSSWH--KLALPQSES------------  872
              :  :||:||||        ||  :  | |: |     ||||||
Db  615  LQFTTKRTNKYAITNKFE-------GTLNISGKVNISMVLPKNESGYDKFKGRTYW  663

Qy  873  KLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGT-----------SGSQ  921
         :|:||  : |||    |    ||  |||   |        |:| :          ||||
Db  664  NLTSLNVSESGEFNLTID-----------SRGSDSAGTLTQPYNLNGISFNKD  705

Qy  922  TVFN-----RLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPT  976
         | ||      :     :| |:  |  |  |:|      |  | :|
Db  706  TTFNVERNARVNFDIKAPIGINKYSSLNYASFNGNISVSGGGSVD----FTLLASSSN  759

Qy  977  MSTPRPIKNAAY----------ATQHGWQGREGLKPLYEMQGALIKQLDAHNVRH  1021
            ||  |:  |           ||  |  :|  |:  :||   ||  :| || || ||
Db  760  VQTPGVVINSKYFNVSTGSSLRFKTSGSTKTGFSIEKDL-TLNATGGNITLLQVEGTDG  817

Qy  1022  NAPQPDLQSKLETLDIGEHGAELLNDMKRFRDELEQSAT----RSVTVLG----QHQ--  1070
          :  : ||:| || | |  |   |:  |  :||   ||::     ||:| |     | :
Db  818  MIGKGIVAKKNITFE---GGNITFGSRKAVTEIEGNVTINNNANVTLIGSDFDNHQKPL  873

Qy  1071  ---------GVLKSNGEI----------NSEFKPSPGKALVQSFNVNRSCQ--------  1102
                    |:||  |:|          |:  |   | | :    :|
Db  874  TIKKDVIINSGNLTAGGNIVNIAGNLTVESNANFK-----AITNFTFNVGGLFDNKGNS  927

Qy  1103  --------DLSKSLQQAVHA-------------------TPPSAESKLQSML  1127
                   |:|||:|  |                      ||  ||:| :|:
Db  928  NISIAKGGARFKDINSKNLSITTNSSSTYRTIISGNITNKNGDLNITNEGSDTEMQ---  984

Qy  1128  GHFVSAGVDMSHQKGEIPL--------------------GRQRDPNDKTALTKSRLILDTVTIGE  1172
          |  ||  ||  || |:||:                     | ||| |:||  :|:| |  ||
Db   985  ----IGGDVSQXEGNLTISSDKINITKQITIKAGVDGENSDSATNNANLTIKTKELKL  1039

Qy  1173  LHEL---ADKAKLVSDHKPDADQIKQLRQOFDTLREKRYESNPVKHYTDMGFTHNKALE  1228
            |:    ||  :  :|    |  :|    :|     |:|  ||   :   | |:| :|
Db  1040  TQDLNISGFNKAEITA--KDGSDLTIGNTNAKKVTFNQVKDSKISADGHKVTLH  1097

Qy  1229  ANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLL--SLDSGESMSFSRS  1286
          :|   :|||  ||  |:  |||  |  |  :|||:||  :|:|      |: ||| |||
Db  1098  S--------KVETSGSNNNTED--SSDNNAGLTIDAKNVTVNNNITSHKAVSISAT  1143

Qy  1287  YGGGVSTVFVPTLSK--KVPVPVIPGA-GITLDRAYNLSFSRTSGGLNVSFGRDGGVS  1341
           |   | |  ||:|    |  | | |  ||| || :|  |  ||||||| || :| |
Db  1144  -SGEITTKTGTTINATTGNVEITAQTGSIIGGIESSSGSVTLTATEGALAVS-----NIS  1197

Qy  1342  GNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSG---TLQGTL  1394
          || ||   |   ||   |:|  |  | | || |: |:| | |  ||   | ||:|
Db  1198  GNTVTVTANSGALTTLAGSTIKGTESVTTSSQ---SGD--IGGTISGGTVEVKATESLTT  1252

Qy  1395  QNSLKFKLTEDE--------LPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTSANL  1449
          |:|  |  |             |  |  |  |  |   :  |  ||    |  | | :| :|
Db  1253  QSNSKIKATTGEANVTSATGTIGGTISGN-------------------TVNVTANA  1289

Qy  1450  -DLRAGINLNEDGSKPNGVTARVSAG--LSASANLAAGSRERSTTSGQFGSTTSASNNR  1505
           || || ||:|  :  ||:|||:||   |||||||||| :| |||||:| |||:|:
Db  1290  GDLTVG-NGAEINATEGAATLTTSSGKLTTEASSHITS-AKGQVNLSAQDGSVAGSINAA  1347

Qy  1506  PTFLNGVGA-------GANLTAALGV---AHSSTHEGKPVGIFPAFTSTNVSAALALDNRT  1556
           | ||  ||       ||| |  | |   | ||  ||     | |:| |  :|| ||
Db  1348  NVTLNTTGTLTTVKGSNINATSGTLVINAKDAELNGAALGNHTVVNATNANGSGSVIATT  1407

Qy  1557  SQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKML--------AALKELD  1602
          |:|:   :   | |:  | |   | |||    |||||      ||| |:|
Db  1408  SSRVNI--TGDLITINGLNIIS-----------KNGINTVLLKGVKIDVKYIQPGIASVDEVI  1457

Qy  1603  DAKPAEQLHILQOHFSAKDVVGDERYEAVRNLKKLVIR  1640
          :||    ||:|  ||||| | ||  | | |||||| ||
Db  1458  EAK-----RILEK--VKD-LSDEEREALAKLGVSAVR  1486
```

```
;  ATTORNEY/AGENT INFORMATION:
;    NAME:  Berkstresser, Jerry W
;    REGISTRATION NUMBER:  22,651
;    REFERENCE/DOCKET NUMBER:  1038-404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-0810
;      TELEFAX:  (703) 415-0813
;  INFORMATION FOR SEQ ID NO: 2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1536 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-302-832-2

Query Match            2.08;   Score 190.5;  DB 1;  Length 1536;
Best Local Similarity  18.6%;  Pred. No. 0.00013;
Matches 297;  Conservative 221;  Mismatches 559;  Indels 521;  Gaps 72;

Qy   304  VTPLAVTLDK---------GKLQLAPDNPPALNTLLKQTLGKDTQHYLAH--HASSD  349
Db   149  INPNGITIGKDAIINTNGFTASTLDISNENIKARNFTFEQTKDKALAEIVNHGLITVGKD  208

Qy   350  GSQHLL---LDNKG---------------HLFDIKSTATSYSVLHNSHPGEIKGKLAQ  389
Db   209  GSVNLIGGKVKNEGVISVNGGSISLLAGQKITISDIINPTITYSIAAPENEAVNLGDIFA  268

Qy   390  AGTGSVSVDG---KSGKISLGSGT-QSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQ  443
Db   269  KG-GNINVRAATIRNQGKLSADSVSKDKSGNIVLSAKEGEA  308

Qy   444  GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQ-----NLSDN  498
Db   309  --------EIGGVISA------QNQQAKGKLMITGDKVTLKTGAVIDLSGK  346

Qy   499  KSSEKIVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGL  558
Db   347  EGGETYL-------GDGERGE-----GKNGIQLAKKTSLEKGSTINVS-------  382

Qy   559  LHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDH  618
Db   383  --GK---EKGGRAIVWGDIALIDGN--------INAQGSGDIAKTGGFV----ETSGHD-  424

Qy   619  QISGFFHDDHGQLNALVKNNPRQQHACPLGNDHQFHPGWNLT-DALVIDNQLGLHHTNPE  677
Db   425  ---LFIKDNAIVDA--------------------KEWLLDFDNVSINAETAGRSNTSE  459

Qy   678  PHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLN  737
Db   460  DDEYTGSGNSASTPKRNK------EKTTLTNTTLES-----------ILKKGTFVNIT  500

Qy   738  INQ----STSSIKHGTENVFSL----------PHVRNKPEPGDALQGLNKDKAQAMAVIGVN  786
Db   501  ANQRIYVNSSINLSNGSLTLWSEGRSGGGVEINNDITTGDDTRGANL--TIYSGGWVDVH  558

Qy   787  KYLALTEKGDI----RSFQIKPGTQOLERPAQTLSREGISGELKDIHVDH---------  832
Db   559  KNISLGAQGNINITAKQDIAFEKGSNQVITGQGTIT----SGNQKGFRFNNVSLNGTGSG  614

Qy   833  -----KQNLYALTHEGEVFHQPREAWQNGAESSSWH---KLALPQSES-----------  872
Db   615  LQFTTKRTNKYAITNKFE----------GTLNLSGKVNISMVLPKNESGYDKFPKGRTYW  663

Qy   873  KLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGT---------SGSQ  921
Db   664  NLTSLNVSESGEFNLTID----------------SRGSDSAGTLTQPYNLNGISFNKD  705

Qy   922  TVFN-----RLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPT  976
Db   706  TTFNVERNARVNFDIKAPIGINKYSSLNYASFNGNISVSGGGSVD-----FTLLASSSN  759

Qy   977  MSTPRPIKNAAY-----------------ATQHGWQGREGLKPLYEMQGALIKQLDAHNVRH  1021
Db   760  VQTPGVVINSKYFNVSTGSSLRFKTSGSTKTGFSIEKDL--TLNATGNITLLQVEGTDG  817

Qy  1022  NAFQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSAT---RSVTVLG----QHQ--  1070
Db   818  MIGKGIVAKKNITFE---GGNITFGSRKAVTEIEGNVTINNNANVLIGSDFDNHQKPL  873

Qy  1071  ------------GVLKSNGEI-----NSEFKPSPGKALVQSFNVNRSGQ---  1102
Db   874  TIKKDVIINSGNLTAGGNIVNIAGNLTVESNANFK-----AITNFTFNVGGLFDNKGNS  927

Qy  1103  ----------DLSKSLQQAVHA------------TPPSAESKLQSML  1127
Db   928  NISTAKGGARFKDIDNSKNLSITTNSSSTYRTIISGNITNKNGDLNITNEGSDTEMQ--  984

Qy  1128  GHFVSAGVDMSHQKGEIPL-----------------GRQRDPNDKTALIKSRLILDTVTIGE  1172
Db   985  -----IGGDVSQKEGNLTISSDKINITKQITIKAGVDGENSDSDATNNANLTKTKELKL  1039

Qy  1173  LHEL---ADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALE  1228
Db  1040  TQDLNISGFNKAETTA--KDGSDLTIGNTNSADGTNAKKVTFNQVKDSKISADGHKVTLH  1097

Qy  1229  ANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLL--SLDSGESMSFSRS  1286
Db  1098  S-----------KVETSGSNNTED--SSDNNAGLTIDAKNVTVNNNTSHKAVSISAT  1143

Qy  1287  YGGGVSTVFVPTLSK--KVPVPVIPGA-GITLDRAYNLSFSRTSGGLNVSFGRDGGVS  1341
Db  1144  -SGEITTKTGTTINATTGNVEITAQTGSILGGIESSSGSVTLTATEGALAVS-----NIS  1197

Qy  1342  GNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSG----TLQGTL  1394
Db  1198  GNTVTVTANSGALTTLAGSTIKGTESVTTSSQ---SGD--IGGTISGGTVEVKATESLTT  1252

Qy  1395  QNSLKFKLTEDE------LPGFIHGLTHGTLIPAELLQKGIEHQMKQGSKLTFSVDTSANL  1449
Db  1253  QSNSKIKATTGEANVTSATGTIGGTISGN----------------TVNVTANA  1289

Qy  1450  -DLRAGINLNEDGSKPNGVTARVSAG---LSASANLAAGSRERSTTSGGFGSTTSASNNR  1505
Db  1290  GDLTVG-NGAEINATEGAATLTTSSGKLTTEASSHTTS-AKGQVNLSAQDGSVAGSINAA  1347

Qy  1506  PTFLNGVGA------GANLTAALGV---AHSSTHEGKPVGIFPAFTSTNVSAALALDNRT  1556
Db  1348  NVTLNTTGTLTTVKGSNINATSGTLVINAKDAELNGAALGNHTVVNATNANGGSVIATT  1407

Qy  1557  SQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKML------AALKELD  1602
Db  1408  SSRVNI--TGDLITINGLNIIS------KNGINTVLLKGVKIDVKYIQPGIASVDEVI  1457

Qy  1603  DAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIR  1640
Db  1458  EAK-----RILEK----VKD-LSDEEREALAKLGVSAVR  1486


RESULT  9
US-08-530-198-2
;  Sequence 2, Application US/08530198
;  Patent No. 5869065
;  GENERAL INFORMATION:
;    APPLICANT:  BARENKAMP, STEPHEN J
;    APPLICANT:  ST. GEME III, JOSEPH W
;    TITLE OF INVENTION:  HIGH MOLECULAR WEIGHT SURFACE PROTEINS
;    TITLE OF INVENTION:  OF NON-TYPEABLE HAEMOPHILUS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
```

us-09-596-784-2.rai

```
;      COUNTRY: U.S.A.
;      ZIP: 22202-0286
;  COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/530,198
;      FILING DATE: 13-DEC-1995
;      CLASSIFICATION: 424
;  ATTORNEY/AGENT INFORMATION:
;      NAME: BERKSTRESSER, JERRY W
;      REGISTRATION NUMBER: 22,651
;      REFERENCE/DOCKET NUMBER: JWB-1186
;  TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (703) 415-0810
;      TELEFAX: (703) 415-0813
;  INFORMATION FOR SEQ ID NO: 2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 1536 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: single
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
US-08-530-198-2
```

```
Query Match           2.0%;   Score 190.5; DB 2;   Length 1536;
Best Local Similarity  18.6%;  Pred. No. 0.00013;
Matches  297;  Conservative 221;  Mismatches 559;   Indels 521;   Gaps   72;
```

```
Qy  304 VTPLAVTLDK----------GKLQLAPDNPPALNTLLKQTLGKDTQHYLAH--HASSD 349
            |:  :        ||    ::| :||  :||: ::|
Db  149 INPNGITIGKDAIINTNGFTASTLDISNENIKARNFTFEQTKDKALAEIVNHGLITVGKD 208

Qy  350 GSQHLL--LDNKG-------------HLFDIKSTATSYSVLHNSHPGEIKGKLAQ 389
            || |  |||            :|| || ::| |  :|
Db  209 GSVNLIGGKVKNEGVISVNGGSISLLAGKITISDIINPTITYSIAAPENEAVNLGDIFA 268

Qy  390 AGTGSVSVDG----KSGKISLGSGT-QSHNKTMLSQPGEAHRSLLTGIMWQHPAGAARPQ 443
            | |:: |       |:| | : : : :  :|| | ||  |  :|
Db  269 KG-GNINVRAATIRNQGKLSADSVSKDKSGNIVLSAKEGEA------------- 308

Qy  444 GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKIYALKDNRTLQ-----NLSDN 498
            |:  |  :: |||  |||    || ||        |   |        ||
Db  309 --------EIGGVISA------QNQQAKGGKLMITGDKVTLKTGAVIDLSGK 346

Qy  499 KSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGL 558
            | ||              ||| ||  |||      || | :   |  |::|
Db  347 EGGETYL---------GGDERGE--------GKNGIQLAKKTSLEKGSTINVS------- 382

Qy  559 LHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDH 618
            ||                 ||    | |:: :|   ::|||  |  ||
Db  383 --GK---EKGGRAIVWGDIALIDGN------INAQGSGDIAKTGGFV-----ETSGHD- 424

Qy  619 QISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT-DALVIDNQLGLHHTNPE 677
            | :|  :| ||  |       :|: :: |  |||       || |:|| ||  |  :|
Db  425 ---LFIKDNAIVDA----------------KEWLLDFDNVSINAETAGRSNTSE 459

Qy  678 PHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLN 737
            | :||  || | :   :|      ||    :: |   ||  | | :| :|
Db  460 DDEYTGSGNSASTPKRNK------EKTTLTNTTLES-----------ILKKGTFVNIT 500

Qy  738 INQ-----STSSIKHGTENVFSL------PHVRNKPEPGDALQLGLNKDDKAQAMAVIGVN 786
            |:|     ::|:|:         |::      |  | :| | |   |::|| :|
Db  501 ANQRIYVNSSINLSNGSLTLWSEGRSGGGVEINNDITTGDDTRGANL--TIYSGGWVDVH 558

Qy  787 KYLALTEKGDI-----RSFQIKPGTQQLERPAQTLSREGISGELKDIHVDH---------- 832
            | ::|::|: |   :|      | |:  ::
Db  559 KNISLGAQGNINITAKQDIAFEKGSNQVITGQGTIT----SGNQKGFRFNNVSLNGTGSG 614

Qy  833 ----KQNLYALTHEGEVFHQPREAWQNGAESSSWH---KLALPQSES------------ 872
```

```
Db      -------------------GTLNISGKVNISMVYLPKNESGYDKFKGRTYW  663
                                 ||:|:::  || ::: ::
Qy  615 LQFTTKRTNKYAITNKFE----------------------------------------

Qy  873 KLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGT--------SGSQ  921
            ||::|||:  : :::  : ||
Db  664 NLTSLNVSESGEFNLTID---------------------SRGSDSAGTLTQPYNLNGISFNKD 705

Qy  922 TVFN-----RLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPT  976
            |  |   :  :| :   |   : :|:|:  :::| :|
Db  706 TTFNVERNARVNFDIKAPIGINKYSSLNYASFNGNISVSGGGSVD-------FTLLASSSN 759

Qy  977 MSTPRPIKNAAY---------ATQHGWQGREGLKPLYEMQGALIKQLDAHNVRH 1021
            ::|::: |::  | |||:|     |:    ::  ||
Db  760 VQTPGVVINSKYFNVSTGSSLRFKTSGSTKTGFSIEKDL--TLNATGNITLLQVEGTDG 817

Qy  1022 NAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSAT------RSVTVLG--QHQ--- 1070
            |:  |  | :| | :|  | || |||   ::|       || ::   : :
Db  818 MIGKGIVAKKNITFE---GNITFGSRKAVTEIEGNVTINNNANVTLIGSDFDNHQKPL  873

Qy  1071 --------GVLKSNGEI------------NSEFKPSPGKALVQSFNVNRSGQ-------- 1102
                    ||:||||                :::
Db  874 TIKKDVIINSGNLTAGGNIVNIAGNLTVESNANFK-----AITNFTNVGGLFDNKGNS  927

Qy  1103 ---------DLSKSLQQAVHA-------------TPPSAESKLQSML 1127
                     :| :|
Db  928 NISIAKGGARFKDIDNSKNLSITTNSSSTYRTIISGNITNKNGDLNITNEGSDTEMQ-- 984

Qy  1128 GHFVSAGVDMSHQKGEIPL---------------GRQRDPNDKTALTKSRLJLDTVIGE 1172
            |  ::|      ::  ||              || : |   ||  :| ||:| ::|
Db  985 ------IGGDVSQKEGNLTISSDKINITKQITIKAGVDGENSDSDATNNANLTIKTKELKL 1039

Qy  1173 LHEL---ADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALE 1228
            | ||   ||:  || ||:   :|:|::  |  :  || :||:||  ||   |:|
Db  1040 TQDLNISGFNKAEITA--KDGSDLTIGNTSADGTNAKKVTFNQVKDSKISADGHKVTLH 1097

Qy  1229 ANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLL--SLDSGESMSFSRS 1286
            ||   |:   |:::| :       || || ::|: ::||: :
Db  1098 S--------------KVETSGSNNNTED--SSDNNAGLTIDAKNVTVNNNITSHKAVSISAT 1143

Qy  1287 YGGGVSTVFVPTLSK--KVPVPVIPGA-GITLDRAYNLSFSRTSGGLNVSFGRDGGVS 1341
            | |    |:|:|:  |||  : |   ::  ||| :: ::  |  |:
Db  1144 --SGEITTKTGTTINATTGNVEITAQTGSILGGIESSSGSVTLTATEGALAVS-----NIS 1197

Qy  1342 GNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSG-------TLQGTL 1394
            ||:|  ||  | |:  | ||   :|  ::  ||  :     ::  |::
Db  1198 GNTVTVTANSGALTTLAGSTIKGTESVTTSSQ---SGD--IGGTISGGTVEVKATESLTT 1252

Qy  1395 QNSLKFKLTEDE----LPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTSANL 1449
            | :|::|     |:  ::   |   |  : |  :|   |
Db  1253 QSNSKIKATTGEANVTSATGTIGGTISGN----------------------TVNVTANA 1289

Qy  1450 -DLRAGINLNEDGSKPNGVTARVSAG---LSASANLAAGSRERSTTSGQFGSTTSASNNR 1505
            ||             |::|:      |:|  || |      ::|:  |::
Db  1290 GDLTVG--NGAEINATEGAATLTTSGKLTTEASSHITS-AKGQVNLSAQDGSVAGSINAA 1347

Qy  1506 PTFLNGVGA------GANLTAALGV---AHSSTHEGKPVGIFPAFTSTNVSAALALDNRT 1556
            |:| ||        |  ||  ||     |::  ::  |  | | :|:: || :
Db  1348 NVTLNTTGLTTVKGSNINATSGTLVINAKDAELNGAALGNHTVVNATNANGSGSVIATT 1407

Qy  1557 SQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKML----------AALKELD 1602
            ||::| ||:  |    |::  | :   |:  |::
Db  1408 SSRVNI--TGDLITINGLNIIS------KNGINTVLLKGVKIDVKYIQPGIASVDEVI 1457

Qy  1603 DAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIR 1640
            ||       :| |  :|  :|::   | || | :|:||
Db  1458 EAK----RILEK--VKD-LSDEEREALAKLGVSAVR 1486
```

```
RESULT 10
US-08-469-880-2
;  Sequence 2, Application US/08469880
;  Patent No. 5876733
;  GENERAL INFORMATION:
```

APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
STREET: Bldg. 1
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

Query Match   2.0%;   Score 190.5; DB 2;   Length 1536;
Best Local Similarity 18.6%; Pred. No. 0.00013;
Matches 297; Conservative 221; Mismatches 559; Indels 521; Gaps 72;

```
Qy  304 VTPLAVTLDK-----------GKLQLAPDNPPALNTLLKQTLGKDTQHYLAH---HASSD 349
            | :| ||                |    ::| | : :||| :|
Db  149 INPNGITIGKDAIINTNGFTASTLDISNENIKARNFTFEQTKDKALAEIVNHGLITVGKD 208

Qy  350 GSQHLL---LDNKG---------HLFDIKSTATSYSVLHNSHPGEIKGKLAQ 389
           || ||    |||            | :|| ||   :||:|| | || ||
Db  209 GSVNLIGGKVKNEGVSVNGGSISLLAGQKITISDIINPTTTYSIAAPENEAVNLGDIFA 268

Qy  390 AGTGSVSVDG---KSGKISLGSGT-QSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQ 443
           | ||:|  ||   | ||| | | |  |::|  :| |   |||
Db  269 KG-GNINVRAATIRNQGKLSADSVSKDKSGNIVLSAKEGEA------------ 308

Qy  444 GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQ-----NLSDN 498
             |  |       |           || :| ||| ||| ::| |      ||||
Db  309 -----EIGGVISA------ONQQAKGGKLMITGDKVTLKTGAVIDLSGK 346

Qy  499 KSSEKLVDKIKSYSVDQRGQVAILDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGL 558
           |::|:| | : ||:|      | |||||
Db  347 EGGETYL---GGDERGE----GKNGIQLAKKTSLEKGSTINVS----------- 382

Qy  559 LHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDH 618
           || :| |  | ::|| | ::|||
```

```
Db  383 -GK---EKGGRAIVWGDIALIDGN------INAQGSGDIAKTGGFV----ETSGHD- 424
          |        | :::| |:| | |        || | ||||| ||
Qy  619 QISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT-DALVIDNQLGLHHTNPF 677

Db  425 ----LFIKDNAIVDA-                    KEWLLDFDNVSINAETAGRSNTSE 459
          | :|| |  :| |                      | : | | ::| ||| | :
Qy  678 PHEILDMGHLGSLALQEGKLHYFDQLRKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLN 737

Db  460 DDEYTGSGNSASTPKRNK----EKTLLTNTTLES----------ILKKGTFVNIT 500
          |      ||:|  | :|     |  | |::|||           :|||||:|:|:
Qy  738 INQ----STSSIKHGTENVFSL----PHVRNKPEPGDALQGLNKDDRAQAMAVIGVN 786

Db  501 ANQRIYVNSSINLSNGSLTLWSEGRSGGGVEINNDITTGDDTRGANL-TIYSGGWVDVH 558
          | | |:  : :||::|| | ::  | :  : :|     ||:|||| | |||  ||:||
Qy  787 KYLALTEKGDI----RSFQIKPGTQQLERPAQTLSREGISGELKDIHVDH-------- 832

Db  559 KNISLGAQGNINITAKQDIAFEKGSNQVITGQGTIT----SGNQKGFRNNVSLNGTGSG 614
          |: || | |::    || ||| | ||:| ||| ||     |||:||| ||:|||| |
Qy  833 -----KQNLYALTHEGEVFHQPREAWQNGAESSSWH--KLALPQSES---------- 872

Db  615 LQFTTKRTNKYAITNKFE------------GTLNISGKVNISMVLPKNESGYDKFKGRTYW 663
          | || ||| |||||| |            |||| | || || :| | :|| | | ||::
Qy  873 KLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGT---------SGSQ 921

Db  664 NLTSLNVSESGEFNLTID--------SRGSDSAGTLTQPYNLNGISFNKD-705
          || || | || | | |         ||||||||| |||| ||| | |:|
Qy  922 TVFN-----RLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPT 976

Db  706 TTFNVERNARVNFDIKAPIGINKYSSLNYASFNGNISVSGGGSVD----FTLLASSSN 759
          |:|| | : | ::|  | || :|| ||| |||:|| || |||||     ||||:|||
Qy  977 MSTPRPIKNAAY--------------ATQHGWQGREGLKPLYEMQGALIKQLDAHNVRH 1021

Db  760 VQTPGVVINSKYFNVSTGGSLRFKTSGSTKTGFSIEKDL--TLNATGGNITLLQVEGTDG 817
          | || || || :|||||||| ||| |||||||| | |   || | |||:|||:| |||
Qy  1022 NAPQPDLQSKLETDLGEHGAELLNDMKRFRDELEQSAT----RSVTVLG----QHQ- 1070

Db  818 MIGKGIVAKKNITFE---GGNITFGSRKAVTEIEGNVTINNNANVTLIGSFDNHQKPL 873
          :| || | |:| |    |||| | | | |||: || ||:|| ||| | |:||:|
Qy  1071 --------------GVLKSNGEI----NSEFKPSPGKALVQSFNVNRSGQ- 1102

Db  874 TIKKDVIINSGNLTAGGNIVNIAGNLTVESNANFK-------AITNFTFNVGGLFDNKGNS 927
          |:|||| :|| |||| |:| ||| ||||:|| ||        ||| | || | |||||:
Qy  1103 -----------------DLSKSLQQAVHA-----TPPSAESKLQSML 1127

Db  928 NISIAKGGARFKDIDNSKNLSITTNSSSTYRTIISGNITNKNGDLNITNEGSDTEMQ- 984
          || || | | | |||:||| |||:|||:|:||: ||:|| ||| || |  | |||
Qy  1128 GHFVSAGVDMSHQKGEIPL------GRQRDPNDKTALTKSRLIIDTVTIGE 1172

Db  985 -----IGGDVSQKEGNLTISSDKINITKQITTIKAGVDGENSDSDATNNANLTIKTKELKL 1039
           | | | |:|||| ||:|||  | ||| |: | ||:|| |: | |:||||| || |||:|
Qy  1173 LHEL----ADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALE 1228

Db  1040 TQDLNISGFNKAEITA--KDGSDLTIGNTNSADGTNAKKVTFNQVKDSKISADGHKVTLH 1097
          | || || | | :|    |||:| | | |:|| |:| | | |:|| || || |||:|||
Qy  1229 ANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLL--SLDSGESMSFSRS 1286

Db  1098 S---------KVETSGSNNNTED--SSDNNAGLTIDAKNVTVNNNITSHKAVSISAT 1143
          |          ||::||:|| |    |:|||| || |:|| :||||| || | ||:|
Qy  1287 YGGGVSTVFVPTLSK--KVPVPVIPGA--GITLDRAYNLSFSRTSGGLNVSFGRDGGVS 1341

Db  1144 -SGEITTKGTTINATTGNVEITAQTGSILGGIESSSGSVTLTATEGALAVS-----NIS 1197
           ||:|| |:| || ||||||:|| ||| || |:|| |||||||||||||      ||:
Qy  1342 GNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSG----TLQGTL 1394

Db  1198 GNTVTVTANSGALTTLAGSTIKGTESVTTSSQ---SGD--IGTISGGTVEVKATESLTT 1252
          || |:|||:| ||||| ||||||||| ||||    |||  || | | ||| |||| |||
Qy  1395 QNSLKFKLTEDE----LPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLITFSVDTSANL 1449

Db  1253 QSNSKIKATTGEANVTSATGTIGGTISGN----------TVNVTANA 1289
          | ||| |||| | |||:| | | ||| ||          |||| ||
Db  1450 -DLRAGINLNEDGSKPNGVTARVSAG--LSASANLAAGSRERSTTSGQFGSTTSASNNR 1505

Qy  1290 GDLLTVG-NGAEINATEGAATLTTSSGKLTFEASSHITS-AKGQVNLSAQDGSVAGSINAA 1347
          ||| | | |||: ||| ||||||||| |||| ||| |:   ||:|||||:| |||| |||
```

```
Qy   350 GSQHLL---LDNKG------------------------------HLFDIKSTATSYSVLHNSHPGEIKGKLAQ 389
         |::|: :|:|
Db   209 GSVNLIGGKVKNEGVISVNGGSISLLAGQKITISDIINPTITYSIAAPENEAVNLGDIFA 268

Qy   390 AGTGSVSVDG---KSGKISLGSGT-QSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQ 443
         ::::  |||  |  ::|::  ||
Db   269 KG-GNINVRAATIRNQGKLSADSVSKDKSGNIVLSAKEGEA------------ 308

Qy   444 GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQ---NLSDN 498
         ||
Db   309 ------------EIGGVISA---QNQQAKGGKLMITGDKVTLKTGAVIDLSGK 346

Qy   499 KSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGL 558
         ::||      :||:    |:|:
Db   347 EGGETYL------GGDERGE---GKNGIQLAKKTSLEKGSTINVS------------- 382

Qy   559 LHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDNELKMKAMPQHALDEHFGHDH 618
         |||       |:||      ||:
Db   383 --GK--EKGGRAIVWGDIALIDGN------INAQGSGDIAKTGGFV---ETSGHD- 424

Qy   619 QISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT-DALVIDNQLGLHHTNPE 677
         ::     :|
Db   425 ----LFIKDNAIVDA-----------------KEWLLDFDNVSINAETAGRSNTSE 459

Qy   678 PHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLN 737
         |:       ||:     ::|                    :||:
Db   460 DDEYTGSGNSASTPKRNK----EKTTLTNTTLES------ILKKGTFVNIT 500

Qy   738 INQ----STSSIKHGTENVFSL---PHVRNKPEPGDALQGLNKDDKAQAMAVIGVN 786
         :||      |::  ||:
Db   501 ANQRIYVNSSINLSNGSLTLWSEGRSGGGVEINNDITTGDDTRGANL-TIYSGGWVDVH 558

Qy   787 KYLALTEKGDI----RSFQIKPGTQQLERPAQTLSREGISGELKDIHVDH--- 832
         |   | |: ::|                           ||:
Db   559 KNISLGAQGNINTAKQDIAFEKGSNQVITGQGTIT----SGNQKGFRNNVSLNGTGSG 614

Qy   833 ------KQNLYALTHEGEVFHQPREAWQNGAESSSWH--KLALPQSES------------ 872
              |::|   :|     |    | :|::|
Db   615 LQFTTKRTNKYAITNKFE------GTLNISGKVNISMVLPKNESGYDKFKGRTYW 663

Qy   873 KLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGT-----SGSQ 921
         :|:|:                        |
Db   664 NLTSLNVSESGEFNLTID------SRGSDSAGTLTQPYNLNGISFNKD 705

Qy   922 TVFN---RLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPT 976
         |.||              |: ::|
Db   706 TTFNVERNARVNFDIKAPIGINKYSSLNYASFNGNISVSGGGSVD----FTLLASSSN 759

Qy   977 MSTPRPIKNAAY------------ATQHGWQGREGLKPLYEMOGALIKQLDAHNVRH 1021
         |:|    ::
Db   760 VQTPGVVINSKYFNVSTGSSLRFKTSGSTKTGFSIEKDL-TLNATGGNITLLQVEGTDG 817

Qy  1022 NAPQPDLQSKLETLDLGEBHGAELLNDMKRFRDELEQSAT---RSVTVLG--QHQ--- 1070
         :  :|                  ||
Db   818 MIGKGIVAKKNITFE---GGNITFGSRKAVTEIEGNVTINNNANVTLIGSDFDNHQKPL 873

Qy  1071 --------GVLKSNGEI-----NSEFKPSPGKALVQSFNVNRSGQ----- 1102
             |::|||:         :|     | :|
Db   874 TIKKDVIINSGNLTAGGNIVNIAGNLTVESNANFK----AITNFTFNVGGLFDNKGNS 927

Qy  1103 ------DLSKSLQQAVHA-----TPPSAESKLQSML 1127
             :|:|:    |            ||:
Db   928 NISIAKGGARFKDIDNSKNLSITTNSSSTYRTIISGNITNKNGDLNITNEGSDTEMQ--- 984

Qy  1128 GHFVSAGVDMSHQKGEIPL----GRQRDPNDKTALTKSRLILDTVTIGE 1172
         | |:|                  ||:
Db   985 --IGGDVSQKEGNLTISSDKINITKQITIKAGVDGENSDSDATNNANLTIKTKELKL 1039

Qy  1173 LHEL----ADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALE 1228
         |:|       :||:|:  |                    ::      :|
Db  1040 TQDLNISGFNKAEITA--KDGSDLTIGNTNSADGTNAKKVTFNQVKDSKISADGHKVTLH 1097
```

```
Qy  1506 PTFLNGVGA-----GANLTAALGV--AHSSTHEGKPVGIFPAFTSTNVSAALALDNRT 1556
         ::|::            |:|||:|     :  |
Db  1348 NVTLNTTGTLTTVKGSNINATSGTLVINAKDAELNGAALGNHTVVNATNANGSGSVIATT 1407

Qy  1557 SQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKML-----AALKELD 1602
         | |::      ::|  ||   ||:|||  |
Db  1408 SSRVNI---TGDLITINGLNIIS------KNGINTVLLKGVKIDVKYIQPGIASVDEVI 1457

Qy  1603 DAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIR 1640
         :||       ||   ||   ||     ||
Db  1458 EAK-----RILEK---VKD-LSDEREALAKLGVSAVR 1486
```

RESULT 11
US-08-728-470-2

```
; Sequence 2, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
;  APPLICANT: Barenkamp, Stephen J
;  TITLE OF INVENTION: High Molecular Weight Surface Proteins
;  TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
;  NUMBER OF SEQUENCES: 10
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Shoemaker and Mattare, Ltd.
;   STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;   STREET: Bldg. 1
;   CITY: Arlington
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 22202-0286
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/728,470
;   FILING DATE:
;   CLASSIFICATION: 424
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/302,832
;   FILING DATE: 16-Mar-1993
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US PCT/US93/02166
;   FILING DATE: 16-Mar-1993
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: GB 9205704.1
;   FILING DATE: 16-Mar-1992
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Berkstresser, Jerry W
;   REGISTRATION NUMBER: 22,651
;   REFERENCE/DOCKET NUMBER: 1038-633
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 415-0810
;   TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 1536 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;  MOLECULE TYPE:
```

US-08-728-470-2

```
Query Match   2.08;  Score 190.5;  DB 2;  Length 1536;
Best Local Similarity  18.6%;  Pred. No. 0.00013;
Matches  297;  Conservative  221;  Mismatches  559;  Indels  521;  Gaps  72;

Qy   304 VTPLAVTLDK--------GKIQLAPDNPPALNTLLKQTLGKDTQHYLAH--HASSD 349
         ::|                :|        |                |
Db   149 INPNGITIGKDAIINTNGTFASTLDISNENIKARNFTFEQTKDKALAEIVNHGLIITVGKD 208
```

```
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1536 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
US-08-617-697-2

Query Match         2.0%;  Score 190.5;  DB 2;  Length 1536;
Best Local Similarity  18.6%;  Pred. No. 0.00013;
Matches  297;  Conservative 221;  Mismatches 559;  Indels 521;  Gaps 72;

Qy   304  VTPLAVTLDK------------GKLQLAPDNPPALNTLLKQTLGKDTQHYLAH--HASSD   349
               :  :::                      :  :  ::  :  :    ::  :
Db   149  INPNGITIGKDAIINTNGFTASTLDISNENIKARNFTEQTKDKALAEIVNHGLITVGKD   208

Qy   350  GSQHLL--LDNKG------------------HLFDIKSTATSYVLHNSHPGEIKGKLAQ   389
          ||  ::  :::::                  :: ::  :  :: :
Db   209  GSVNLIGGKVKNEGVISVNGGSISLLAGKITISDIINPITYSIAAPENEAVNLGDIFA   268

Qy   390  AGTGSVSVDG---KSGKISLGSGT-QSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQ   443
          :  :: ::      :::  : ::   :::   :  :     :::
Db   269  KG-GNINVRAATIRNQGKLSADSVSKDKSGNIVLSAKEGEA------------------   308

Qy   444  GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQ------NLSDN   498
               :                :               ::    :
Db   309  ------------------EIGGVISA-------QNQQAKGGKLMITGDKVTLKTGAVIDLSGK   346

Qy   499  KSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGL   558
           :  :  :               ::   :                ::     :  :
Db   347  EGGETYL-----GGDERGE-------GKNGIQLAKKTSLEKGSTINVS   382

Qy   559  LHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDH   618
           ||                 :  :     :                            :::
Db   383  -GK--EKGGRAIVWGDIALIDGN------INAQGSGDIAKTGGFV-----ETSGHD-   424

Qy   619  QISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT-DALVIDNQLGLHHTNPE   677
              ::   :  :  :                          :
Db   425  ---LFIKDNAIVDA----------KEWLLDFDNVSINAETAGRSNTSE   459

Qy   678  PHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAEESDCKQLKKGLDGAAYLLKDGEVKRLN   737
           : :: :          :: : :      : :  :    ||:
Db   460  DDEYTGSGNSASTPKRNK----EKTTLTNTTLES--------ILKKGTFVNIT   500

Qy   738  INQ-------STSSIKHGTENVFSL----PHVRNKPEPGDALQGLNKDKAQAMAVIGVN   786
           :         ::  ::   :  ::      :      :  :
Db   501  ANQRIYVNSSINLSNGSLTIWSEGRSGGGVEINNDITTGDDTRGANL-TIYSGGWVDVH   558

Qy   787  KYLALTEKGDI---------RSFQIKPGTQQLERPAQTLSREGISGELKDIHVDH------   832
           :  :   ::           ::  ::                      ::
Db   559  KNISLGAQGNINTTAKQDIAFEKGSNQVITGQGTIT---SGNQKGFRFNNVSLNGTGSG   614

Qy   833  --------KQNLYALTHEGEVFHQPREAWQNGAESSSWH---KLALPQSES---------   872
                  :::          :                 :  :
Db   615  LQFTTKRTNKYAITNKFE----------GTLNISGKVNISMVLPKNESGYDKFKGRTYW   663

Qy   873  KLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGT----SGSQ   921
           | ::: :   :    :        :  ::               ||  ::
Db   664  NLTSLNVSESGEFNLTID----SRGSDSAGTLTQPYNLNGISFNKD   705

Qy   922  TVFN----RLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPT   976
           ||:     :         :    :   :      :
Db   706  TTFNVERNARVNFDIKAPIGINKYSSLNYASFNGNISVSGGGSVD---FTLLASSSN   759

Qy   977  MSTPRPIKNAAY--------------ATOHGWQGREGLKPLYEMQGALIKQLDAHNVRH   1021
            :: :  :  :               :      : :  :
Db   760  VQTPGVVINSKYFNVSTGSSLRFKTSGSTKTGFSIEKDL--TLNATGGNITLLQVEGTDG   817

Qy   1022 NAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSAT---RSVTVLG---QHQ---   1070
           : :   : : :  : : :   ||:   : :               :
Db   818  MIGKGIVAKKNITFE----GNITFGSRKAVTEIEGNVTINNNANVTLIGSDFDNHQKPL   873

Qy   1229 ANYDAVKAFINAFKKEHHGVNLTRTVLESQGSAELAKKLKNTLL--SLDSGESMSFSRS   1286
           :                   ||    :      :::  :  ::: :  :::
Db   1098 S------------KVETSGSNNNTED--SSDNNAGLTIDAKNVTVNNNTSHKAVSISAT   1143

Qy   1287 YGGGVSTVFVPTLSK---KVPVPVIPGA-GITTLDRAYNLSFSRTSGGLNVSFGRDGGVS   1341
           :  :: :   : ::     : :  :    :    : : :  :     :   |||    :
Db   1144 -SGEITTKTGTTINATTGNVEITAQTGSILGGIESSSGSSVTLTATEGALAVS-----NIS   1197

Qy   1342 GNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSG------TLQGTL   1394
           ||  : :            ::     : :  :  : ::   : :          ::
Db   1198 GNTVTVTANSGALTTLAGSTIKGTESVTTSSQ---SGD--IGGTISGGTVEVKATESLTT   1252

Qy   1395 QNSLKFKLTEDE------LPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTSANL   1449
           :::  :    ::         :             :                : :
Db   1253 QSNSKIKATTGEANVTSATGTIGGTISGN------------TVNVTANA   1289

Qy   1450 -DLRAGINLNEDGSKPNGVTARVSAG---LSASANLAAGSRERSTTSGQFGSTTSASNNR   1505
            ||    :  ::       ||   ::    ||      ||   || :      :: :  :
Db   1290 GDLTVG-NGAEINATEGAATLITSSGKLTTEASSHITS-AKGQVNLSAQDGSVAGSINAA   1347

Qy   1506 PTFLNGVGA------GANLTAALGV-----AHSSTHEGKPVGIFPAFTSTNVSAALALDNRT   1556
            :    | :       ||:   |          : : :           ||:   : |:|
Db   1348 NVTLNTTGTLTTVKGSNINATSGTLVINAKDAELNGAALGNHTVVNATNANGSGSVIATT   1407

Qy   1557 SQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKML--------------AALKELD   1602
           ||:           ::   : : ::  :                        ||:  ::
Db   1408 SSRVNI---TGDLITINGLNIIS------KNGINTVLLKGVKIDVKYIQPGIASVDEVI   1457

Qy   1603 DAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIR   1640
           ||                 :   :            :
Db   1458 EAK---RILEK---VKD-LSDEREALAKLGVSAVR   1486
```

RESULT 12
US-08-617-697-2
; Sequence 2, Application US/08617697
; Patent No. 5977336
; GENERAL INFORMATION:
;   APPLICANT: Barenkamp, Stephen J
;   TITLE OF INVENTION: High Molecular Weight Surface Proteins
;   TITLE OF INVENTION: of No. 5977336 Typeable Haemophilus
;   NUMBER OF SEQUENCES: 11
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Shoemaker and Mattare, Ltd.
;   STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;   STREET: Bldg. 1
;   CITY: Arlington
;   STATE: Virginia
;   COUNTRY: U.S.A.
;   ZIP: 22202-0286
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/617,697
;   FILING DATE: 01-APR-1996
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/302,832
;   FILING DATE: 05-OCT-1994
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US PCT/US93/02166
;   FILING DATE: 16-MAR-1993
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Berkstresser, Jerry W
;   REGISTRATION NUMBER: 22,651
;   REFERENCE/DOCKET NUMBER: 1038-557
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (703) 415-0810
;   TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:

```
;   FILING DATE:     01-APR-1996
;   CLASSIFICATION:     424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/302,832
;     FILING DATE:    05-OCT-1994
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US PCT/US93/02166
;     FILING DATE:    16-MAR-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Berkstresser, Jerry W
;     REGISTRATION NUMBER:  22,651
;     REFERENCE/DOCKET NUMBER:  1038-557
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (703) 415-0810
;     TELEFAX:    (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1600 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;  US-08-617-697-10
```

```
Query Match         2.0%;      Score 187;   DB 2;   Length 1600;
Best Local Similarity   19.1%;   Pred. No. 0.00024;
Matches  328;  Conservative  234;  Mismatches  632;  Indels  520;  Gaps  82;

Qy    304  VTPLAVTLDK----------------GKLQLAPDNPPALNTLLKQTLGKDTQHYLAH--HASSD  349
               :  :|  :|                  :|| :| ||| |:| : | ||:|:|  :|:      |: ||
Db    149  INPNGITIGKDAIINTNGFTASTLDISNENIKARNFTLEQTKDKALAEIVNHGLITVGKD  208

Qy    350  GSQHLL---LDNKG---------------------HLFDIKSTATSYSVLHNSHP------GEIKG  385
               ||::|       ||::|                     || |:| | :| |:  |:      | :|
Db    209  GSVNLIGGKVKNEGVISVNGGSISLLAGQKITISDIINPTITYSIAAPENEAINLGDIFA  268

Qy    386  KLAQAGTGSVSVDGKSGKISLGSGT--QSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQ  443
               :| :|| | ||: || :||:|:|      || |  ||: ::| :|||:|   | |  ||
Db    269  KGGNIVRAATIRNK-GKLSADSVSKDKSGNIVLSAKEGEA-----------------------  308

Qy    444  GESIRLHDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQ--------NLSDN  498
               |  :| |  || :|  :|   || | :| |  ::|   |:|  ||  |        :| | ||
Db    309  ----------EIGGVISA------------QNQQAKGKLMITGDKVTLKTGAVIDLSGK  346

Qy    499  KSSEKLVDKIKSYSVDQRGQVAILTDPGRHKMSIMPSLDASPESHISLSLHFADAHQGL  558
               :  | | :|| |:| | : |: ||   ::||::|  |:|:||  ||| :| | ||  ::|
Db    347  EGGETYL-------GDERGE-------GKNGIQLAKKTLEKGSTINVS---------------  382

Qy    559  LHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDH  618
               :|   |       :| |: :|   | |  :| || :||| : :: :|    | | || |   |||
Db    383  --GK----EKGGRAIVWGDIALIDGNINAQGSDIAKTG------------------GFVETSGHDL  424

Qy    619  QISG--------FFHDDHGQLNALV--KNNFRQOHACPLGNDHQFHPGWNLTDALVIDN  667
               : |:|        || | || | |    ||| |  |  ||  || |||:| :| |:| |:
Db    425  SIGDDVIVDAKEWLLDPDDVSIETLTSGRNNTGENOGYTTGDGTKESPKGNSISKPTL--  482

Qy    668  QLGLHHTNPEPHEILDMGHL-------------GSLALQEGKLHYFDQLTKGWTGAESDCK  715
               :|| |    | : | ||                    |  :| |: |    ||:  ||  | |
Db    483  -------TNSTLEQILRRGSYVNITANNRIYVNSSINLSNGSL--------------------  518

Qy    716  QLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHV--RN-------KPEP  764
               :|: || | |: ||: || :: :|:|     |  | |::|        ::       :|
Db    519  -------------TLHTKRDG----VKINGDITSNENGNLTIKAGSWVDVHKNITLGTFLNIVA  566

Qy    765  GDALQGLNKDDKAQ----AMAVIGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTL  816
               | | |:| |  |      | |:| :| |  :|| |:  | |:|   :
Db    567  GDSVAFERGDKARNATDAQITAQGTITVNK----------DDKQFR----------------F  604

Qy    817  SREGISGELKDIHVDHKQNLYALTHEGEV-------FHQPRE--AWQNGAESSSWHKLA  866
               :| |: | :|: || ||| | | | ||           ||||      | | :|| :||
Db    605  NNVSLNGTGKGLKFIANQNNFTHKFDGEINISGIVTINQTTKKDVKYWNASKDSYWNVSS  664

Qy    867  LPQSESKLKSLDMSHEHKPIATFEDGSQHOLKAGGWHAYAAPERGPLAVGTSGSQTVFNR  926
```

```
Qy   1071  ------------GVLKSNGEI-------------NSEFKPSPGKALVQSFNVNRSCQ--------  1102
                           |  :|  | :|                :|  : |
Db    874  TIKKDVIINSGNLTAGGNIVNIAGNLTVESNANFK------AITNFTENVGGLFDNKGNS  927

Qy   1103  ------------DLSKSLQQAVHA------------------------TPPSAESKLQSML  1127
                           : |  :| |  :|                         :::||
Db    928  NISIAKGGARFKDIDNSKNLSITTNSSSTYRTIISGNITNKNGDLNITNEGSDTEMQ---  984

Qy   1128  GHFVSAGVDMSHQKGEIPL----------------GRQRDPNDKTALTKSRLLDTVTIGE  1172
                | :  |:   |:  |: |                :| |  :|: |:  ::  :| |
Db    985  ---IGGDVSQKEGNLTISSDKINITKQITIKAGVDGENSDSDATNNANLTIKTKELKL  1039

Qy   1173  LHEL---ADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALE  1228
                || |     |  | | |  |        |:|: :|  | | :  :|  : :| | ||
Db   1040  TQDLNISGFNKAEITA-KDGSDLTIGNTNSADGTNAKKVTFNQVKDSKISADGHKVTLH  1097

Qy   1229  ANYDAVKAFINAFKKEHHGVNLTRTVLESQGSAELAKKLKNTLL--SLDSGESMSFSRS  1286
                | :  |   | |    |  |:||  |:: | |   | :|::|       |||  | |  ||
Db   1098  S-----------KVETSGSNNNTED--SSDNNAGLTIDAKNVTVNNITSHKAVSISAT  1143

Qy   1287  YGGGVSTVFVPTLSK--KVPVPIPGA-GITLDRAYNLSFSRITSGGLNVSFGRDGGVS  1341
                || | :|  |  |:|  |||: | |   | ::| :||  |:|||:| | :||  | ||
Db   1144  -SGEITTKTGTTINATTGNVEITAQTGSILGGIESSSGSVTLTATEGALAVS-----NIS  1197

Qy   1342  GNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSG-------TLQGTL  1394
                ||  :|      | || :|::|   || :|| :|:|   | |: :|        |:  ||
Db   1198  GNTVTVTANSGALTLAGS|IKGTESVTTSSQ---SGD--IGGTISGGTVEVKAFSLTT  1252

Qy   1395  QNSLKFKLTEDE-------LPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTSANL  1449
                ::|     |  |           |:|  |:|::| |    :|::| | | :|  ::: |: ||
Db   1253  QSNSKIKATTGEANVTSATGTIGGTISGN-----------------TVNVTANA  1289

Qy   1450  -DLRAGINLNEDGSKPNGVTARVSAG---LSASANLAAGSRERSTTSGQFGSTTSASNNR  1505
                   | |  | :|| | | |   | | |    |  |: :: |  | |||  |  || | | ||
Db   1290  GDLTVG-NGAEINATEGAATLTTSSGKLTTEASSHITS-AKGQVNLSAQDGSVAGSINAA  1347

Qy   1506  PTFLNGVGA-------GANLTAALGV----AHSSTHEGKPVGIFPAFTSTNVSAALALDNRT  1556
                   :| | |        | :|  | |      ||: | | :|  ||   | :| |||  |: |
Db   1348  NVTLNTTGTLTTVKGSNINATSGTLVINAKDAELNGAALGNHTVVNATNANGSGSVIATT  1407

Qy   1557  SQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKML------------AALKELD  1602
                   :| ||  | | |:||  ||  || |  | |::|               | :|: |
Db   1408  SSRVNI----TGDLITINGLNIIS------KNGINTVLLKGVKIDVKYIQPGIASVDEVI  1457

Qy   1603  DAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIR  1640
                :||    |:|    | | |:| | |   |:  :|  |:|
Db   1458  EAK-----RILEK----VKD-LSDEEREALAKLGVSAVR  1486
```

```
RESULT 13
US-08-617-697-10
;  Sequence 10, Application US/08617697
;  Patent No. 5977336
;  GENERAL INFORMATION:
;    APPLICANT:  Barenkamp, Stephen J
;    TITLE OF INVENTION:  High Molecular Weight Surface Proteins
;    TITLE OF INVENTION:  of No. 5977336-Typeable Haemophilus
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Shoemaker and Mattare, Ltd.
;      STREET:  2001 Jefferson Davis Hwy., 1203 Crystal Plaza
;      STREET:  Bldg. 1
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202-0286
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/617,697
```

```
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
;   APPLICANT: Zhu, Xueliang
;   APPLICANT: Lee, Wen-Hwa
;   TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Campbell and Flores
;     STREET: 4370 La Jolla Village Drive, Suite 700
;     CITY: San Diego
;     STATE: California
;     COUNTRY: USA
;     ZIP: 92122
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/328,254
;     FILING DATE: 24-OCT-1994
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/141,239
;     FILING DATE: 22-OCT-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Campbell, Cathryn A.
;     REGISTRATION NUMBER: 31,815
;     REFERENCE/DOCKET NUMBER: P-CJ 1191
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (619) 535-9001
;     TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2482 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-328-254-6

Query Match                 2.0%;  Score 186.5;  DB 1;  Length 2482;
Best Local Similarity   19.3%;  Pred. No. 0.00053;
Matches  288;  Conservative  245;  Mismatches  592;  Indels  369;  Gaps  75;

Qy   488 DNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISL 547
             ||.  ||.:  :||||::: |||:||  :||.|         ||
Db     4 EKENLQSKINHLETCLKTQQIKSHEYNER--------VRTLEMDRE--NL 43

Qy   548 SLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGK--------LFSAAIPKQGDGN 598
Db    44 SVEIRNLHNVLDSKSVEVETQKLAYMELQQKAEFSDQKHQKEIENMCLKTSQLTGQVEDL 103

Qy   599 ELKMKAMPGHALDEHFGHDHQISGFFHDHGQLNALVKNNFRQQHACPLGNDHQFHPGWN 658
Db   104 EHKLQLLSNEIMDKDRCYQD----LHAEYESLRDLLKS---KDASLVTNEDHQ----- 149

Qy   659 LTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHY-FDQLTKGWTGAESDCKQL 717
Db   150 -RSLLAFDQQPAMIHSFAN----IIGEQGSMPSERSECRLEADQSPKNSAILQNRVDSL 203

Qy   718 KKGLDGAAYLLKD---------GEVKR---- 749
Db   204 EFSLESQKQMNSDLQKQCEELVQIKGEIEENLMKAEQMHQSFVAETSQRISKLQEDTSAH 263

Qy   750 ENVF--SLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQ 807
Db   264 QNVVAETLSALENKEK-----ELQLLNDKVETEQAEIQELKKSNHLLE-DSLKELQLLSETL 319

Qy   808 QLE---------RPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFH--------- 847
```

```
Db   665 L------------TLNTVQKFTFIKFVDSGSNGQDLRSSRRSFAGVHFNGI-----GGKTNFN- 710
Qy   927 LMQGVKGKVI------PGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPT 976
Db   711 --IGANAKALFLKPNAATDPKKELPITFNANTTA-TGNSDSSVMFDIHANLTSRAAGIN 767
Qy   977 MSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNA--PQPDLQSKLET 1034
Db   768 MDS----------INITGGLDFSITSHNRNSNAFEIKKDL-----T 798
Qy  1035 LDLGEHGAELLNDMKRFRDELEQSATRS-----VTVLGQHQGVLKSNGEINSEFKPSPGKAL 1091
Db   799 INATGSNFSLKQTKDSFYNEYSKHAINSSHNLTILG---GNVTLGGE-NSSSSITGNINI 854
Qy  1092 VQSFNV-------NRSGQDLSK-------SLQQAVHATPPSAE 1120
Db   855 TNKANVTLQADFSNSNTGLKKRTLTLGNISVEGNLSLTGANANIVGNLSIAEDSTFKGEA 914
Qy  1121 SKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTAL--------TKSRLILDVT--I 1170
Db   915 SDNLNITGTFTNNGTANINIKGVVKLG--DINNKGGLNITTNASGTQKTIINGNITNEK 971
Qy  1171 GELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEAN 1230
Db   972 GDLN-----IKNIKADA------EIQIGGNISQKEGNLTISSDKYNITN 1009
Qy  1231 YDAVKAFINAFK---KEHHGVNLTTRT-VLESQGSAELAKKLKNTLLSLDSGESMSFSRS 1286
Db  1010 QITIKAGVEGGRSDSSEAENALTIQTKELKLAGDLNIS-GFNKAEITAKNGSDLITGNA 1068
Qy  1287 YGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFS---RTSGGLNVSFGRDGGVSGN 1343
Db  1069 SGGNAD------AKKVTFDKVKDSKISTD-GHNVTLNSEVKTSNG-SSNAGNDNS-TGL 1118
Qy  1344 IMVATGHDVMPYMTGKKT----SAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQN-S 1397
Db  1119 TISAKDVTVNNNVTSHKTINISAAAGN-----VTTKEGTTIN----ATTGSVEVTAQNGT 1169
Qy  1398 LKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTSANLDLRAGI-- 1455
Db  1170 IKGNITSQNV--------TVTATENLVT-TENAVINATSGTVNISTKTG-DIKGGIES 1217
Qy  1456 ---NLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGV 1512
Db  1218 TSGNVNITAS--GNTLKVSNITGQDVTVTADAGALTTTA--GSTISATT------- 1262
Qy  1513 GAGANLTRALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSISLELKRAEPVTS 1572
Db  1263 -GNANITTKTGDIN---GKVESSSGSVTVATGATLAVGNISGNTVTI---------T 1307
Qy  1573 NDISELTSTLGKHFK--DSATTKMLAALKELDDAKPAEQLHLLQQHFSAKDV----VGD 1625
Db  1308 ADSGKLTSVGSTINGTNSVTT----------SSQSGDIEGTISGNTVNVTASTGD 1353
Qy  1626 ERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPA 1685
Db  1354 --LTIGNSAKVEARNGAA-TLTAESGKL--TTQTGSSITSSNGQTTLTAK--DSSI-A 1403
Qy  1686 SSAKRLGEMMNNDPALKDI-IKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVGREEV 1744
Db  1404 GNINAANVTLNTTGTLTTTGDSKINAT---SGTLTINAKD-------AKLDGAAS--- 1448
Qy  1745 GVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERN-IGTINFKYGQDQ 1803
Db  1449 ----GDRTVVNATNASGSGNVTAK-----TSSSVNITGDLNTINGLNIISENGR 1493
Qy  1804 NTPRRFTLEGGIAQANPQVASALTDLKKEGLEMK 1837
Db  1494 NTVVRLRGKEIDVKYIQPGVASV-----EEVIRAK 1522
```

RESULT 14

```
RESULT 15
US-08-353-700-1
; Sequence 1, Application US/08333700
; Patent No. 5599919
; GENERAL INFORMATION:
;   APPLICANT: YEN, TIMOTHY J.
;   APPLICANT: RATTNER, JEROME B.
;   TITLE OF INVENTION: NUCLEIC ACID ENCODING A
;   TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
;   TITLE OF INVENTION: AND METHODS OF USE
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
;     STREET: 1601 MARKET STREET, SUITE 720
;     CITY: PHILADELPHIA
;     STATE: PA
;     COUNTRY: USA
;     ZIP: 19103-2307
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/353,700
;     FILING DATE: 09-DEC-1994
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: REED, JANET E.
;     REGISTRATION NUMBER: 36,252
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (215) 563-4100
;     TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 3248 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;     ORGANISM: HUMAN
US-08-353-700-1


Query Match           2.0%;  Score 184.5;  DB 1;  Length 3248;
Best Local Similarity  18.4%;  Pred. No. 0.0011;
Matches 363;  Conservative 292;  Mismatches 759;  Indels 563;  Gaps 89;

Qy 108 RDLLAR------DDGETQHEAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQ 162
        ||   |      :|
Db 340 RDELVRTTAQYDQASTKYTALEQKLKLKKLTEDLSCQRQNAESAR------CSLEQKIKEK 392

Qy 163 QKR------HQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGST 216
        :|:      ::|  |  |
Db 393 EKEFQEELSRQQRSFQTLDQECIQMKA--------RLTQELQQAKNMHNVLQAELDKLT 443

Qy 217 SKATTAHADRVEIAQE-------DDDSEFQQLHQQRLARERENPPOPPKLGVATPISARFQPK 272
        :| | ::  :|
Db 444 SVKQQLENNLEEFKQKLCRAEQAFQASQIKENELRRSMEEMKKENNL-----LKSHSEQK 498

Qy 273 LTAVAESVLEGTDTTOSPLKPQSMLKGSGAGVTPLJ-----------AVTILDKGKL 316
        ||      :     :::|
Db 499 AREVCHLEAELKNIKQCLNQSONFAEEMKAKNTSQETMLRDLQEKINQQENSLTLEKLKL 558

Qy 317 QLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLH 376
        :|         :  ::
Db 559 AVA------DLEKQ-----------RDCSQDLLKKREHHI 581
```

```
Db 320 SLEKKEMSSIISLNKREIEELTQE--NGTLKEINASLNQEKMNLIQKSESFANYIDEREK 377

Qy 848 ------------QPREAWQNGAESSSWHKLALPQSESKLKSLSLDMSHEHKPIATFE 890

Db 378 SISELSDSDQYKQEKLILLQRCEETGNAYEDLSQKYKAAQEKNSKLECL--LNECTSLCENR 435

Qy 891 DGSQHQLKAGGWHAYAAPER--GPLAVGTSGSQTVFNRL--MQGVKGKVIPGSGLTVK 944

Db 436 KNELEQLK----EAFAKEHQEFLTKLAFAEERNQNLMLELETVQQALRSEMTDNQN---N 488

Qy 945 LSAQTGGM------TGAEGRKVSSKFSERIRA-YAFNPTMST-------PRPIKNAAYA 989

Db 489 SKSEAGGLKQEIMTLKEEQNKMQKEVNDLLQENEQLMKVMKTKHECQNLESEPIRNSV-K 547

Qy 990 TQHGWQGREGLKPLYEMQGALIKQ-LDAHNVRHNAPQPDLQSKLETLDLGEHGAELL-N 1046

Db 548 ERESERNQCNFKPQMDLE---VKEISLDSYNAQLVQLEAMLRNKELKLQESEKEKECLQH 604

Qy 1047 DMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPG-KALVQSFNVNRSGQDLS 1105

Db 605 ELQTIRGDLETSNLQDMQ-SQEISGLK-DCEIDAEEKYISGPHELSTSQNDN--AHLQ 658

Qy 1106 KSLQ---------QAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDP 1152

Db 659 CSLQTTMNKLNELEKICEILQAEKYELVTELNDSRSECITATRKMAEEVGKL-LNEVKIL 717

Qy 1153 NDKTALTKSRLILDTVTIGELHELADKAKLVS--------DHKPDADQIKQLRQQFDT 1202

Db 718 NDDSGLLHGELVED-IPGGEFGEQPNEQHVSLAPLDESNSYEHLTLSD--KEVQMHFAE 774

Qy 1203 LREKRYESNPVKHYTDMGFTHNK-ALEANYDAVKAFINAFKKEHHGVNLTTRTVLES-Q 1259

Db 775 LQEK-FLSLQSEH---KILHDQHCQMSSKMSELQTYVDSLKAE---NLVLSTNLRNFQ 825

Qy 1260 GSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAY 1319

Db 826 G--DLVKEMQ---LGLEEG----------LVPSLSSSC-VPDSSSLSSLGDSSF 863

Qy 1320 NLSFSRTSGGLNVSFGRDGGVSGN-----IMVATGHDVMPYMTGKKTSAGNASD----1368

Db 864 YRALLEQTGDMSLLSNLEGAVSANQCSVDEVFCSSLQE-ENLTRKETPSAPAKGVEELE 921

Qy 1369 -----WLSAKHKISPDLRIGAAVSGTLQGTLQNS---------------LKFKLTE 1404

Db 922 SLCEVYRQSLEKLEEKME-------SQGIMKNKEIQELEQLLSSERQELDCLRKQYLSE 973

Qy 1405 DELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTSANLDLRAGINLN----- 1458

Db 974 NE--OWQQKLTSVTLE---MESKLAAEKKQTEQLSLELEV-ARLQLQ-GLDLSSRSLLG 1025

Qy 1459 ---EDGSKPNGVTARVSAGLSASA----------------NLAAGSRERSTTSGQ 1494

Db 1026 IDTEDAIQGRNESCDISKEHTSETTERTPKHDVHQICDKDAQQDLNL--DIEKITETGA 1082

Qy 1495 FGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAT'STNVSAALALDN 1554

Db 1083 LKPTGECSGEQSPDTNYEPPGEDKTQGSSECISEL-----------SFSGPNALVPMDFLG 1132

Qy 1555 RTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQ 1614

Db 1133 NQEDIHNLQLRVKE--TSNENLRLLHVI-EDRDRKVESLLNEMKELD-----SKLHLQE 1183

Qy 1615 QHFSAKDVVGDERYEAVRNLKKIV-IRQQAAD-SHSMELGSASHSTTYNNLSRI-NNDG 1670

Db 1184 VQLMTK------IEACIELEKIVGELKKENSDLSEKLEYFSCDHQEL--LQRVETSEG 1233

Qy 1671 IVELLHKHFDAALPASSAKRLGEMMN-ND---PALKDIIKQLQSTPFSSASVSME---L 1722

Db 1234 LNSDLEMHAD----KSSREDIGDNVAKVNDSWKERFLDVENELSRIRSEKASIEHEALYL 1289

Qy 1723 KDGLR-EQTEKAILD-----GKV---GREEVGVLFQDRNNLRVKSVSVSQSVS 1766

Db 1290 EADLEVVQTEKICLEKDNENKQKVIVCLEEELSVVTSERNQLRGELDTMSKKTT 1343
```

```
Qy    377  NSHPGEIKGKLAQAGTGSVSVDGKSGKISLSGSTQSHNKTMLSQPGEAHRSLLTGIWQHP   436
Db    582  ----EQLNDKL--------------------------SKTEKESKALLS----          600

Qy    437  AGAARPQGESIRLHDDKIHILHPE---LGVWQSADKDTHSQLSRQADGKLYALKDNRTLQ   493
Db    601  -ALELKKKEYEELKEEKTLFSCWKSENEKLLTQMESE------KEN--LQ            641

Qy    494  NLSDNKSSEKLVDKIKISYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFAD 553
Db    642  SKINHLETCLKTQQIKSHEYNER-----------VRTLEMDRE---NLSVEIRN        681

Qy    554  AHQGLLHGKSELEAQSVAISHGRLVVADSEGK--------LFSAAIPKQGDNELKMKA     604
Db    682  LHNVLDSKSVEVETQKLAYMELQQKAEFSDQKHQKEIENMCLKTSQLTGQVEDLEHKLQL  741

Qy    605  MPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALV  664
Db    742  LSNEIMDKDRCYQD----LHAEYESLRDLLKS---KDASLVTNEDHQ------RSLLA    786

Qy    665  IDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHY-FDQLTKGWTGAESDCKQLKKGLDG  723
Db    787  FDQQPAMHHSFAN-----IIGEQGSMPSERSECRLEADQSPKNSAILQNRVDSLEFSLES  841

Qy    724  AAYLLKD-----------GEVKR---------LNINQSTSSIKHGT---ENVF-       753
Db    842  QKQMNSDLQKQCEELVQIKGEIEENLMKAEQMHQSFVAETSQRISKLQEDTSAHQNVVAE  901

Qy    754  SLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLE---  810
Db    902  TLSALENEK---ELQLLNDKVETEQAEIQELKKSNHLLE-DSLKELQLLSETLSLEKKE   957

Qy    811  -RPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFH-----------            847
Db    958  MSSIISLNKREIEELTQE--NGTLKEINASLNQEKMNLIQKSESFANYIDEREKSISELS 1015

Db    848  -------QPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQ      896
Db   1016  DQYKQEKLILLQRCEETGNAYEDLSQKYKAAQEKNSKLECL--LNECTSLCNRKNELEQ  1073

Qy    897  LKAGGWHAYAAPER--GPLAVGTSGSQTVFNRL---MQGVKGKVIPGSGLTVKLSAQTG   950
Db   1074  LK----EAFAKEHQEFLTKLAFAEERNQNLMLELETVQQALRSEMTDNQN---NSKSEAG 1126

Qy    951  GM-----TGAEGRKVSSKFSERIRA-YAFNPTMST--------PRPIKNAAYATQHGWQ   995
Db   1127  GLKQEIMTLKEEQNKMQKEVNDLLQENEQLMKVMKTKHECQNLESEPIRNSV-KERESER 1185

Qy    996  GREGLKPLYEMQGALIKQ--LDAHNVRHNAPQPDLQSKLETLDLGEHGAELL-NDMKRFR 1052
Db   1186  NQCNFKPQMDLE--VKEISLDSYNAQLVQLEAMLRNKELKIQESEKEKECLQHELQTIR  1242

Qy   1053  DELEQSATRSVTVLGQHQGVLIKSNGEINSEFKPSPG-KALVQSFNVNRSGQDLSKSLQ- 1109
Db   1243  GDLETSNLQDMQ-SQEISGLK-DCEIDAEEKYISGPHELSTSQNDN---AHLQCSLQTT  1296

Qy   1110  --------QAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTAL    1158
Db   1297  MNKLNELEKICEILQAEKYELVTELNDSRSECITATRKMAEEVGKL-LNEVKILNDDSGL 1355

Qy   1159  TKSRLILDTVTIGELHELADKAKLVS---------DHKPDADQIKQLRQQFDTLREK--  1206
Db   1356  LHGELEVED-IPGGEFGEQPNEQHPVSLAPLDESNSYEHLTLSD--KEYQMHFAELQEKFL 1412

Qy   1207  ----------RYESNPVKHYHTDMGFTHNKALEANYDAVKAFINAFKKEHH----     1246
Db   1413  SLQSEHKILHDQHCQMSSKMSELQTYVDSLKAENLVLSTN--LRNFQGDLVKEMQLGLE  1469

Qy   1247  -----------GVNLTTRTVLESQGSAELAKKLKNTL---LSLD-----          1276
Db   1470  EGLVPSLSSSCVPDSSSLSSLGDSSFYRALLEQTGDMSLLSNLEGAVSANQCSVDEVFCS 1529

Qy   1277  ---------SGESMSFS---RSYGG------GVSTVFVPTLSKKVPVPIPGAGITLD   1316
```

```
Db   1530  SLQTYVDSLKAENLVLSTNLRNFQGDLVKEMQLGLEEGLVPSLSSSC-VPDSSSLSSLGD 1588

Qy   1317  RAYNLSFSRTSGGLNVSFGRDGGVSGN-----IMVATGHDVMPYMTGKKTSAGNASD-- 1368
Db   1589  SSFYRALLEQTGDMSLLSNLEGVVSANQCSVDEVFCSSLQE--ENLTRKETPSAPAKGVE 1646

Qy   1369  --------WLSAKHKISPDLRIGAAVSGTLQGTLQNS---------LKFK        1401
Db   1647  ELESLCEVYRQSLEKLEKME------SQGIMKNKEIQELEQLLSEROELDCLRKQY     1698

Qy   1402  LTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVDTSANLDLRAGINLN-   1458
Db   1699  LSENE--QWQQKLTSVTLE--MESKLAAEKKQTEQLSLELEV-ARLQLQ-GLDLSSRS  1750

Qy   1459  ------EDGSKPNGVTARVSAGLSASA------NLAAGSRERSTT               1491
Db   1751  LLGIDTEDAIQGRNESCDISKEHTSETTERTPKHDVHQICDKDAQQDLNL--DIEKITE 1807

Qy   1492  SGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALA 1551
Db   1808  TGAVKPTGECSGEQSPDTNYEPPGEDKTQGSSECISEL-------SFSGPNALVPMD   1857

Qy   1552  LDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLH 1611
Db   1858  FLGNQEDIHNLQLRVKE-TSNENLRLLHVI-EDRDRKVESLLNEMKELD------SKLH 1908

Qy   1612  ILQQHFSAKDVVGDERYEAVRNLKKLV-IRQQAAD-SHSMELGSASHSTTYNNLSRI-N 1667
Db   1909  LQEVQLMTK-----TEACIELEKIVGELKKENSDLSEKLEYFSCDHQEL--LQRVET  1958

Qy   1668  NDGIVELLHKHFDAALPASSAKRLGEMMN-ND---PALKDIIKQLQSTPFSSASVSME- 1721
Db   1959  SEGLNSDLEMHAD------KSSREDIGDNVAKVNDSWKERFLDVENELSRIRSEKASIEHEA 2014

Qy   1722  --LKDGLR-EQTEKAILD------GKV--GREEVGVLFQDRNNLRVKSVSVQSVS    1766
Db   2015  LYLEADLEVVQTEKLCLEKDNENKQKVIVCLEEELSVVTSERNQLRGELDTMSKKTT   2071
```

Search completed: June 5, 2001, 18:16:59
Job time: 185 sec

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:       June  5, 2001, 18:13:14 ; Search time 55.68 Seconds
                                         (without alignments)
                                         1886.962 Million cell updates/sec

Title:        US-09-596-784-2
Perfect score: 9448
Sequence:     1 MELKSLGTEHKAAVHTAAHN.........NPQVASALTDLKKEGLEMKS 1838

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:     390729 seqs, 5163235 residues

Total number of hits satisfying chosen parameters:       390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    A_Geneseq_0401:*
        1:    /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
        2:    /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
        3:    /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
        4:    /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
        5:    /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
        6:    /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
        7:    /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
        8:    /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
        9:    /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
       10:    /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
       11:    /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
       12:    /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
       13:    /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
       14:    /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
       15:    /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
       16:    /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
       17:    /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
       18:    /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
       19:    /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
       20:    /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
       21:    /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*
       22:    /SIDS6/gcgdata/geneseq/geneseqp/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Score	Length	DB	ID	Description
1	9448	100.0	9448	1838	20	W98011	Hypersensitive res
2	9448	100.0	9448	1838	21	Y71095	Erwinia amylovora
3	9448	100.0	9448	1838	21	X84856	A hypersensitive r
4	1041	11.0	201	201	21	Y69258	Amino acid sequenc
5	275	2.9	2514	2514	21	Y75097	Neisseria meningit
6	245	2.6	2599	2599	21	Y75098	Neisseria meningit
7	241.5	2.6	1837	1837	19	Y85564	Human homologue of
8	232	2.5	1981	1981	21	W42634	Protein sequence t
9	225	2.4	2608	2608	21	Y85574	Hs-UNC-53/3/GFP f
10	224.5	2.4	2541	2541	21	B41087	Human ORFX ORF851
11	222	2.3	2385	2385	21	Y85569	Human homologue of
12	221.5		2048	2048	21	Y75096	Neisseria gonorrhe
13	210	2.2	3647	3647	11	R05041	Filamentous haemag
14	207.5	2.2	3596	3596	21	Y87407	Bordetella pertuss
15	204.5	2.2	1992	1992	17	W04505	Moraxella 200 kDa
16	203.5	2.2	1780	1780	19	W53863	Human gravin polyp
17	203.5	2.2	1780	1780	21	B15380	Human gravin prote
18	203.5	2.1	1978	1978	20	Y27230	Amino acid sequenc
19	201	2.1	2353	2353	17	R99393	Haemophilus adhesi
20	198	2.1	1968	1968	12	R10941	Mutant protease (d
21	197.5	2.1	1962	1962	21	B29694	Mouse FLASH protei
22	197	2.1	1962	1962	12	R10558	Mutant protease (A
23	197	2.1	2272	2272	18	W21731	GAL4/HA/NuMA fusio
24	195	2.1	2326	2326	21	Y71158	Rat phosphodiester
25	194	2.1	2042	2042	19	W56319	Haemophilus paraga
26	194	2.0	2954	2954	20	Y01632	Amino acid sequenc
27	192	2.0	1974	1974	12	R10940	Mutant protease (d
28	191	2.0	1962	1962	12	R10561	Mutant protease (N
29	191	2.0	1962	1962	12	R10557	Mutant protease (A
30	191	2.0	2192	2192	18	W21732	LexA/NuMA fusion p
31	190.5	2.0	1536	1536	14	R41723	High molecular wei
32	190	2.0	2411	2411	21	B23860	Haemophilus influe
33	189.5	2.0	1536	1536	14	R41725	High molecular wei
34	189.5	2.0	1536	1536	18	W30293	Non-typeable Haemo
35	189.5	2.0	2115	2115	21	Y49937	Human NuMA protein
36	189	2.0	1931	1931	21	B23258	Human apoptosis in
37	188.5	2.0	1420	1420	20	W81025	AIB1 (Amplified in
38	188.5	2.0	1522	1522	20	Y21975	Human steroid rece
39	188.5	2.0	1536	1536	15	R63505	Haemophilus high m
40	188.5	2.0	1536	1536	21	B01846	Haemophilus influe
41	188.5	2.0	1962	1962	12	R10560	Mutant protease (K
42	188.5	2.0	1962	1962	12	R10563	Mutant protease (K
43	188.5	2.0	2039	2039	19	W56322	Haemophilus paraga
44	188.5	2.0	2442	2442	21	Y77575	Human cytoskeletal
45	188.5	2.0	2517	2517	21	Y71159	Human phosphodiest

ALIGNMENTS

```
RESULT   1
W98011
ID   W98011 standard; Protein; 1838 AA.
XX
AC   W98011;
XX
DT   21-JUN-1999  (first entry)
XX
DE   Hypersensitive response elicitor DspE.
XX
KW   Hypersensitive response elicitor; DspE; disease resistance;
KW   insect resistance; biological control; transgenic plant.
XX
OS   Erwinia amylovora.
XX
PN   WO9907206-A1.
XX
PD   18-FEB-1999.
XX
PF   24-JUL-1998;  98WO-US15426.
XX
PR   06-AUG-1997;  97US-0055105.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Beer SV,  Bogdanove AJ,  Kim JF,  Wei Z;
XX
DR   WPI; 1999-180362/15.
DR   N-PSDB; X24810.
XX
PT   Nucleic acid encoding hypersensitive response-eliciting protein -
PT   used to improve growth of plants and impart resistance to disease
PT   and insects
XX
```

```
PS   Claim 18; Page 51-56; 75pp; English.
XX
CC   This polypeptide comprises the 198 kDa hypersensitive response
CC   elicitor protein DspE of Erwinia amylovora. The nucleotide
CC   sequence of the dsp region of E. amylovora strain Ea321 was
CC   determined using subclones of pCPP430. A two-gene operon was
CC   discovered comprising dspE (see X24810) and dspF (see X24811). The
CC   isolated dsp DNA molecules and encoded proteins can be used to
CC   impart disease resistance to plants, to enhance plant growth,
CC   and/or to control insects on plants. This is achieved by applying
CC   a hypersensitive response elicitor protein or polypeptide in a
CC   non-infectious form to plants or plant seeds, or by producing
CC   transgenic plants or plant seeds transformed with DNA encoding a
CC   hypersensitive response elicitor. Protection can be provided
CC   against a wide range of viruses, bacteria, fungi and insects, e.g.
CC   tobacco mosaic virus and tomato mosaic virus, Pseudomonas syringae,
CC   Xanthomonas campestris, Fusarium oxysporum, Phytophthora infestans,
CC   armyworm, diamondback moth, etc. The method avoids use of
CC   infectious agents or polluting chemicals. Claimed transgenic
CC   plants are selected from alfalfa, rice, wheat, barley, rye, cotton,
CC   sunflower, peanut, corn, potato, bean, pea, chicory, lettuce,
CC   endive, cabbage, brussel sprout, sweet potato, beet, parsnip,
CC   turnip, cauliflower, broccoli, turnip, radish, spinach, onion,
CC   garlic, eggplant, pepper, celery, carrot, squash, pumpkin,
CC   zucchini, cucumber, apple, pear, melon, citrus, strawberry, grape,
CC   raspberry, pineapple, soybean, tobacco, tomato, sorghum, sugarcane,
CC   Arabidopsis thaliana, Sainpaulia, petunia, pelargonium, poinsettia,
CC   chrysanthemum, carnation and zinnia.
XX
SQ   Sequence   1838 AA;

Query Match       100.0%;    Score 9448;    DB 20;    Length 1838;
Best Local Similarity 100.0%;    Pred. No. 0;
Matches 1838;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

Qy      1  MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQP    60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  melkslgtehkaavhtaahnpvghgvalqqgssssspqnaaaslaaegknrgkmprihqp    60

Qy     61  STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQH   120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  staadgisaahqqkksfslrgclgtkkfsrsapqgqpgtthskgatlrdllarddgetqh   120

Qy    121  EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS   180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  eaaapdaarltrsggvkrrnmddmagrpmvkggsgedkvptqqkrhqlnnfgqmrqtmls   180

Qy    181  KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ   240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  kmahpasanagdrlqhspphipgshheikeepvgstskattahadrveiaqeddddsefqq   240

Qy    241  LHQQRLARERENPQQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS   300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  lhqqrlarerenpqqppklgvatpisarfqpkltavaesvlegtdttqsplkpqsmlkgs   300

Qy    301  GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG   360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  gagvtplavtldkgklqlapdnppalntllkqtlgkdtqhylahhassdgsqhllldnkg   360

Qy    361  HLFDIKSTATSYSVLINSHPGEIKGKLAQAGTGVSVVDGKSGKISLGSGTQSHNKTMLSQ   420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  hlfdikstatsysvlinshpgeikgklaqagtgvsvvdgksgkislgsgtqshnktmlsq   420

Qy    421  PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD   480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  pgeahrslltgiwqhpagaarpqgesirlhddkihilhpelgvwqsadkdthsqlsrqad   480

Qy    481  GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGVAILTDTPGRHKMSIMPSLDAS   540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  gklyalkdnrtlqnlsdnksseklvdkiksysvdqrgqvailtdtpgrhkmsimpsldas   540

Qy    541  PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL   600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  peshislslhfadahqgllhgkseleaqsvaishgrlvvadsegklfsaaipkqgdgnel   600

Qy    601  KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT   660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  kmkampqhaldehfghdhqisgffhddhgqlnalvknnfrqqhacplgndhqfhpgwnlt   660

Qy    661  DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG   720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  dalvidnqlglhhtnpepheildmghlgslalqegklhyfdqltkgwtgaesdckqlkkg   720

Qy    721  LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM   780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  ldgaayllkdgevkrlninqstssikhgtenvfslphvrnkpepgdalqglnkddkaqam   780

Qy    781  AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT   840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  avigvnkylaltekgdirsfqikpgtqqlerpaqtlsregisgelkdihvdhkqnlyalt   840

Qy    841  HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG   900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  hegevfhqpreawqngaessswhklalpqseseklksldmshehkpiatfedgsqhqlkag   900

Qy    901  GWHAYAAPERGPLAVGTSGSGQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  gwhayaapergplavgtsgsgqtvfnrlmqgvkgkvipgsgltvklsaqtggmtgaegrkv  960

Qy    961  SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  sskfseriryayafnptmstprpiknaayatqhgwqgreglkplyemqgalikqldahnvr  1020

Qy   1021  HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN  1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  hnapqpdlqskletldlgehgaellndmkrfrdeleqsatrsvtvlgqhqgvlksngein  1080

Qy   1081  SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ  1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  sefkpspgkalvqsfnvnrsgqdlskslqqavhatppsaesklqsmlghfvsagvdmshq  1140

Qy   1141  KGEIPLGRQRDPNDKTALTKSRLILDDVTIGELHELADKAKLVSDHKPDADQIKQLRQQF  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  kgeiplgrqrdpndktaltksrlildvtigelheladkaklvsdhkpdadqikqlrqqf  1200

Qy   1201  DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG  1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  dtlrekryesnpvkhytdmgfthnkaleanydavkafinafkkehhgvnlttrtvlesqg  1260

Qy   1261  SAELAKKLKNTLLSLDSGESMSFSRSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  saelakklkntllsldsgesmsfsrsyggvstvfvptlskkvpvpvipgagitldrayn  1320

Qy   1321  LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHHISPDL  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  lsfsrtsgglnvsfgrdggvsgnimvatghdvmpymtgkktsagnasdwlsakhhispdl  1380

Qy   1381  RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  rigaavsgtlqgtlqnslkfkltedelpgfihgltngtltpaellqkgiehqmkqgsklt  1440

Qy   1441  FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  fsvdtsanldlraginlnedgskpngvtarvsaglsasanlaagsrersttsggftstts  1500

Qy   1501  ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI  1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  asnnrptflngvgaganltaalgvahssthegkpvgifpaftstnvsaalaldnrtsqsi  1560

Qy   1561  SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK  1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  slelkraepvtsndiseltstlgkhfkdsattkmlaalkelddakpaeqlhilqqhfsak  1620

Qy   1621  DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD  1680
```

```
Db   1621 dvvgderyeavrnlkklvirqqadshsmelgsashsttynnlsrinndgivellhkhfd 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 aalpassakrlgemmnndpalkdiikqlqstpfssasvsmelkdglreqtekailddgkvg 1740

Qy   1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLGTSNSAAMSMERNIGTINFKYG 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 reevgvlfqdrnnlrvksvsvsqsvsksegfntpalllgtsnsaamsmernigtinfkyg 1800

Qy   1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
          ||||||||||||||||||||||||||||||||||||||
Db   1801 qdqntprrftleggiaqanpqvasaltdlkkeglemks 1838
```

RESULT   2
Y71095
ID   .Y71095 standard; Protein; 1838 AA.
XX
AC   Y71095;
XX
DT   08-SEP-2000 (first entry)
XX
DE   Erwinia amylovora hypersensitive response elicitor encoded by dspE gene.
XX
KW   Hypersensitive response elicitor; environmental stress resistance;
KW   plant; pathogen; dspE gene.
XX
OS   Erwinia amylovora.
XX
PN   WO20002805S-A2.
XX
PD   18-MAY-2000.
XX
PF   04-NOV-1999; 99WO-US26039.
XX
PR   05-NOV-1998; 98US-0107243.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z, Schading RL;
XX
DR   WPI; 2000-376566/32.
DR   N-PSDB; D00670.
XX
PT   Application of a hypersensitive response elicitor protein to plants to
PT   impart stress resistance -
XX
PS   Disclosure; Page 15-20; 84pp; English.
XX
CC   The patent discloses a method to impart stress resistance to plants by
CC   applying a hypersensitive response elicitor in a non-infectious form to
CC   a plant or seed. The present sequence is a hypersensitive response
CC   elicitor encoded by dspE gene from Erwinia amylovora.
CC   The protein elicits plant pathogen's hypersensitive response and is
CC   used to impart stress resistance to plants.
XX
SQ   Sequence 1838 AA;

```
Query Match      100.0%;   Score 9448;  DB 21;  Length 1838;
Best Local Similarity  100.0%;  Pred. No.0;
Matches 1838;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy     1 MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAAASLAAEGKNRGKMPRIHQP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 melkslgtehkaavhtaahnpvghgvalqqgssssspqnaaaaslaaegknrgkmprihqp 60

Qy    61 STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQH 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 staadgisaahqqkksfslrgclgtkkfsrsapqgqpgtthskgatlrdllarddgetqh 120

Qy   121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 eaaapdaarltrsggvkrrnmddmagrpmvkggsgedkvptqqkrhqlnnfgqmrqtmls 180

Qy   181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 kmahpasanagdrlqhspphipgshheikeepvgstskattahadrveiaqeddddsefgq 240

Qy   241 LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 lhqqrlarerenppqppkplgvatpisarfqpkltavaesvlegtdttqsplkpqsmlkgs 300

Qy   301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKG 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 gagvtplavtldkgkiqlapdnppalntllkqtlgkdtqhylahhassdgsqhllldnkg 360

Qy   361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 hlfdikstatsysvlhnshpgeikgklaqagtgsvsvdgksgkislgsgtqshnktmlsq 420

Qy   421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 pgeahrslltgiwqhpagaarpqgesirlhddkihilhpelgvwqsadkdthsqlsrqad 480

Qy   481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLLDAS 540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 gklyalkdnrtlqnlsdnksseklvdkiksysvdqrgqvailtdtpgrhkmsimpsldas 540

Qy   541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL 600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 peshislslhfadahqgllhgkseleaqsvaishgrlvvadsegklfsaaipkqgdgnel 600

Qy   601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 kmkampqhaldehfghdhqisgffhddhgqlnalvknnfrqqhacplgndhqfhpgwnlt 660

Qy   661 DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKOLKKG 720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 dalvidnqlglhhtnpepheildmghlgslalqegklhyfdqltkgwtgaesdckqlkkg 720

Qy   721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQLNKDDKAQAM 780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 ldgaayllkdgevkrlninqstssikhgtenvfslphvrnkpepgdalqlnkddkaqam 780

Qy   781 AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT 840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 avigvnkylaltekgdirsfqikpgtqqlerpaqtlsregisgelkdihvdhkqnlyalt 840

Qy   841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHOLKAG 900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 hegevfhqpreawqngaessswhklalpqseskiksldmshehkpiatfedgsqhqlkag 900

Qy   901 GWHAYAAPERGPLAVGTSGSQTVFNRLMOGVKGKVIPGSGLTVKLSAQTGMTGAEGRKV 960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 gwhayaapergplavgtsgsqtvfnrlmqgvkgkvipgsgltvklsaqtggmtgaegrkv 960

Qy   961 SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 sskfserirayafnptmstprpiknaayatqhgwqgreglkplyemqgalikqldahnvr 1020

Qy  1021 HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 hnapqpdlqskletldlgehgaellndmkrfrdeleqsatrsvtvlgqhqgvlksngein 1080

Qy  1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 sefkpspgkalvqsfnvnrsggdlskslqqavhatpppsaesklqsmlghfvsagvdmshq 1140

Qy  1141 KGEIPLGRQRDPNDKTALTKSRLILDVTIGELHELADAKLVSDHKPDADQIKQLRQQF 1200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 kgeiplgrqrdpndktaltksrlildvtigelheladaklvsdhkpdadqikqlrqqf 1200

Qy  1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
XX
PT   Hypersensitive response elicitor polypeptides useful for imparting
PT   enhanced growth, disease resistance and insect resistance to plants,
PT   especially vegetables and ornamental flowers -
XX
PS   Disclosure; Page 17-22; 100pp; English.
XX
CC   The present sequence represents a hypersensitive response elicitor
CC   polypeptide. The polynucleotide represents the dspE gene. The
CC   specification describes hypersensitive response elicitor polypeptide
CC   fragments, which do not elicit a hypersensitive response. Instead,
CC   the proteins impart disease resistance to plants, enhance plant
CC   growth, and/or control insects. The polypeptide fragments may be
CC   used to these properties to plants. The plants which may be treated
CC   in this way include vegetables, crops and ornamental plants such as
CC   alfalfa, rice, wheat, barley, rye, cotton, sunflower, peanut, corn,
CC   potato, sweet potato, bean, pea, chicory, lettuce, endive, cabbage,
CC   brussel sprout, beet, parsnip, turnip, cauliflower, broccoli, radish,
CC   spinach, onion, garlic, eggplant, pepper, celery, carrot, squash,
CC   pumpkin, zucchini, cucumber, apple, pear, melon, citrus, strawberry,
CC   grape, raspberry, pineapple, soybean, tomato, tobacco, sorghum or
CC   sugarcane, Arabidopsis thaliana, Saintpaulia, petunia, pelargonium,
CC   poinsettia, chrysanthemum, carnation or zinnia.
XX
SQ   Sequence  1838 AA;
XX

Query Match             100.0%;  Score 9448;  DB 21;  Length 1838;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1838;       Conservative  0;      Mismatches  0;     Indels  0;   Gaps  0;

Qy      1 MELKSLGTEHKAAVHTAAHNPVGHGVALQOGSSSSPQNAAASLAAEGKNRGKMPRIHQP 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 melkslgtehkaavhtaahnpvghgvalqogsssspqhaaaslaaegknrgkmprihqp 60

Qy     61 STAADGISAAHQKKSFSLRGCLGTKKFSRSAPQOGPGTTHSKGATLRDLLARDDGETQH 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 staadgisaahqkksfslrgclgtkkfsrsapqggpgtthskgatlrdllarddgetqh 120

Qy    121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 eaaapdaarltrsggvkrrnmddmagrpmvkggsgedkvptqqkrhqlnnfgqmrqtmls 180

Qy    181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEPVGTSKATTAHADRVEIAQEDDDSEFQQ 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 kmahpasanagdrlqhspphipgshheikeepvgtskattahadrveiaqeddsefqq 240

Qy    241 LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 lhqqrlarerenppqppkigvatpisarfqpkltavaesvlegtdttqsplkpqsmlkgs 300

Qy    301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSOHLLLDNKG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gagvtplavtldkgklqlapdnppalntllkqtlgkdtqhylahhassdgsghllldnkg 360

Qy    361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 hlfdikstatsysvlhnshpgeikgklaqagtgsvsvdgksgkislgsgtqshnktmlsq 420

Qy    421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 pgeahrslltgiwqhpagaarpqgesirlhddkihilhpelgvwqsadkdthsqlsrqad 480

Qy    481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 gklyalkdnrtlqnlsdnksseklvdkiksysvdqrgqvailtdtpgrhkmsimpsldas 540

Qy    541 PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 peshislslhfadahqgllhgkseleaqsvaishgrlvvadsegklfsaaipkqgdgnel 600

Qy    601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1201 dtlrekryesnpvkhytdmgfthnkaleanydavkafinafkkehhgvnltrtvlesgg 1260
Qy   1261 SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 saelakklkntllsldsgesmsfsrsygggvstvfvptlskkvpvpvipgagitldrayn 1320
Qy   1321 LSFSRTSGGLNVSFGRDGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 lsfsrtsgglnvsfgrdggvsgnimvatghdvmpymtgkktsagnasdwlsakhkispdl 1380
Qy   1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 rigaavsgtlqnslkfkltedelpgfihgltThgtltpaellqkgiehqmkqggsklt 1440
Qy   1441 FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 fsvdtsanldlraginlnedgskpngvtarvsaglsasanlaagsrersttsgqfgstts 1500
Qy   1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 asnnrptflngvgaganltaalgvahssthegkpvgifpaftstnvsaalaldnrtsqsi 1560
Qy   1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 slelkraepvtsndiseltstlgkhfkdsattkmlaalkelddakpaeqlhilqqhfsak 1620
Qy   1621 DVVGDERYEAVRNLKKLVIRQQAADSHMELGSASHSTTYNNLSRINNDGIVELLHKHFPD 1680
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 dvvgderyeavrnlkklvirqqaadshmelgsashsttynnlsrinndgivellhkhfd 1680
Qy   1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 aalpassakrlgemmnndpalkdiikqlqstpfssasvsmelkdglreqtekaildgkvg 1740
Qy   1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLGTSNSAAMSMERNIGTINFKYG 1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 reevgvlfqdrnnlrvksvsvsqsvsksegfntpallgtsnsaamsmernigtinfkyg 1800
Qy   1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
          ||||||||||||||||||||||||||||||||||||||
Db   1801 qdqntprrftleggiaqanpqvasaltdlkkeglemks 1838
```

```
RESULT  3
Y84856
ID   Y84856 standard; Protein; 1838 AA.
XX
AC   Y84856;
XX
DT   08-AUG-2000  (first entry)
XX
DE   A hypersensitive response elicitor protein.
XX
KW   Hypersensitive response; insect control; disease resistance;
KW   hypersensitive response elicitor; plant growth; vegetable; crop;
KW   ornamental plant; dspE gene.
XX
OS   Erwinia amylovora.
XX
PN   WO20002010452-A2.
XX
PD   13-APR-2000.
XX
PF   05-OCT-1999;  99WO-US23181.
XX
PR   05-OCT-1998;  98US-0103050.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z,  Fan H,  Niggemeyer JL;
XX
DR   WPI; 2000-303745/26.
DR   N-PSDB; A14940.
```

Qy 1741 REEVGVLFQDRNNLRVKSVSVSQSVKSEGFNTPALLIGTSNSAAMSMERNIGTINFKYG 1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1741 reevgvlfqdrnnlrvksvsvsqsvksegfntpalligtsnsaamsmernigtinfkyg 1800

Qy 1801 QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
         ||||||||||||||||||||||||||||||||||||||
Db 1801 qdqntprrftleggiaqanpqvasaltdlkkeglemks 1838

RESULT 4
Y69258
ID Y69258 standard; Protein; 201 AA.
XX
AC Y69258;
XX
DT 30-MAY-2000 (first entry)
XX
DE Amino acid sequence of the secretion signal of the DspE protein.
XX
KW Type III secretion system; DspE protein; hairpin secretion system;
KW effector protein; Avr protein; avirulence protein; agriculture.
XX
OS Erwinia amylovora.
XX
PN WO200002996-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999;   99WO-US15425.
XX
PR 10-JUL-1998;   98US-0092357.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Bauer DW,   Beer SV,   Bogdanove AJ,   Collmer A,   Ham JH;
XX
DR WPI; 2000-182218/16.
XX
PT New DNA construct encoding type III secretion system, used for
PT recombinant production of secreted protein and for identifying
PT potential effector proteins -
XX
PS Claim 6; Page 14-15; 71pp; English.
XX
CC The present sequence represents a functional type III secretion
CC system from the N-terminal domain of the DspE protein. The DspE
CC protein is secreted by a type III secretion system. The DspE secretion
CC signal is compatible with the hairpin secretion system of Erwinia
CC amylovora. The secretion signal polynucleotide is used to produce
CC the constructs of the invention, which also comprise a promoter and a
CC second DNA that encodes a protein or polypeptide that can be secreted
CC by the type III secretion signal. Host cells containing the constructs
CC are used for recombinant production of the polypeptide. The constructs
CC are also used to screen for potential effector proteins, e.g. Avr
CC (avirulence) proteins potentially useful in agriculture.
XX
SQ Sequence   201 AA;

Query Match   11.0%;   Score 1041;   DB 21;   Length 201;
Best Local Similarity   100.0%;   Pred. No. 3.2e-58;
Matches 201;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

Qy   1 MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQP 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 melkslgtehkaavhtaahnpvghgvalqqgsssspqnaaaslaaegknrgkmprihqp 60

Qy 61 STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQH 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 staadgisaahqqkksfslrgclgtkkfsrsapqgqpgtthskgatlrdllarddgetqh 120

Qy 121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGOMRQTMLS 180

Db   601 kmkampqhaldehfghdhqisgffhddhgqlnalvknnfrqqhacplgndhgfhpgwnlt 660

Qy   661 DALVIDNQLGLHHTNPEPHEILDMGHGLHLIGSLALQEGKLHYFDQLTKGWTGAEESDCKQLKKG 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 dalvidnqlglhhtnpepheildmghglhligslalqegklhyfdqltkgwtgaeesdckqlkkg 720

Qy   721 LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 ldgaayllkdgevkrlningstssikhgtenvfslphvrnkpepgdalqglnkddkaqam 780

Qy   781 AVIGVNKYIALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 avigvnkyialtekgdirsfqikpgtqqlerpaqtlsregisgelkdihvdhkqnlyalt 840

Qy   841 HEGEVFHQPREAWQNGAESSSWHKLALPQSESKIKSLDMSHEHKPIATFEDGSQHQLKAG 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 hegevfhqpreawqngaessswhklalpqseskiksldmshehkpiatfedgsqhqlkag 900

Qy   901 GWHAYAAAPERGPLAVGTSGSQOTVFNRLMQGVKGKVIPGSGLTVKLSAQTGMTGAEGRKV 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 gwhayaaapergplavgtsgsqtvfnrlmqgvkgkvipgsgltvklsaqtgmtgaegrkv 960

Qy   961 SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 sskfserirayafnptmstprpiknaayatqhgwqgreglkplyenqgalikqldahnvr 1020

Qy 1021 HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 hnapqpdlqskletldlgehgaellndmkrfrdeleqsatrsvtvlgqhqgvlksngein 1080

Qy 1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 sefkpspgkalvqsfnvnrsgqdlskslqqavhatppsaesklqsmlghfvsagvdmshq 1140

Qy 1141 KGEIPLGRQRDPNDKTALTKSRLLDTVTIGELHELADRAKLVSDHKPDADQIKQLRQQF 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 kgeiplgrqrdpndktaltksrlldtvtigelheladrakalvsdhkpdadqikqlrqqf 1200

Qy 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 dtlrekryesnpvkhytdmgfthnkaleanydavkafinafkkehhgvnlttrtvlesqg 1260

Qy 1261 SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYN 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 saelakklkntllsldsgesmsfsrsyggvstvfvptlskkvpvpvipgagitldrayn 1320

Qy 1321 LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 lsfsrtsgglnvsfgrdggvsgnimvatghdvmpymtgkktsagnasdwlsakhkispdl 1380

Qy 1381 RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLT 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 rigaavsgtlqgtlqnslkfkltedelpgfihgltthgtltpaellqkgiehqmkqgsklt 1440

Qy 1441 FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1441 fsvdtsanldlraginlnedgskpngvtarvsaglsasanlaagsrersttsgqfgstts 1500

Qy 1501 ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1501 asnnrptflngvgaganltaalgvahssthegkpvgifpaftstnvsaalaldnrtsqsi 1560

Qy 1561 SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1561 slelkraepvtsndiseltstlgkhfkdsattkmlaalkelddakpaeqlhilqqhfsak 1620

Qy 1621 DVVGDERYEAVRNLKLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1621 dvvgderyeavrnlklklvirqqaadshsmelgsashsttynnlsrinndgivellhkhfd 1680

Qy 1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1681 aalpassakrlgemmnndpalkdiikqlqstpfssasvsmelkdglreqtekaildgkvg 1740

Db   1741 reevgvlfqdrnnlrvksvsvsqsvksegfntpalligtsnsaamsmernigtinfkyg 1800

Db   1801 qdqntprrftleggiaqanpqvasaltdlkkeglemks 1838

Db   601 kmkampqhaldehfghdhqisgffhddhgqlnalvknnfrqqhacplgndhgfhpgwnlt 660

Db   661 dalvidnqlglhhtnpepheildmghglhligslalqegklhyfdqltkgwtgaeesdckqlkkg 720

Db   721 ldgaayllkdgevkrlningstssikhgtenvfslphvrnkpepgdalqglnkddkaqam 780

Db   781 avigvnkyialtekgdirsfqikpgtqqlerpaqtlsregisgelkdihvdhkqnlyalt 840

Db   841 hegevfhqpreawqngaessswhklalpqseskiksldmshehkpiatfedgsqhqlkag 900

Db   901 gwhayaaapergplavgtsgsqtvfnrlmqgvkgkvipgsgltvklsaqtgmtgaegrkv 960

Db   961 sskfserirayafnptmstprpiknaayatqhgwqgreglkplyenqgalikqldahnvr 1020

Db   1021 hnapqpdlqskletldlgehgaellndmkrfrdeleqsatrsvtvlgqhqgvlksngein 1080

Db   1081 sefkpspgkalvqsfnvnrsgqdlsksklqqavhatppsaesklqsmlghfvsagvdmshq 1140

Db   1141 kgeiplgrqrdpndktaltksrlldtvtigelheladkaklvsdhkpdadqikqlrqqf 1200

Db   1201 dtlrekryesnpvkhytdmgfthnkaleanydavkafinafkkehhgvnlttrtvlesqg 1260

Db   1261 saelakklkntllsldsgesmsfsrsyggvstvfvptlskkvpvpvipgagitldrayn 1320

Db   1321 lsfsrtsgglnvsfgrdggvsgnimvatghdvmpymtgkktsagnasdwlsakhkispdl 1380

Db   1381 rigaavsgtlqgtlqnslkfkltedelpgfihgltthgtltpaellqkgiehqmkqgsklt 1440

Db   1441 fsvdtsanldlraginlnedgskpngvtarvsaglsasanlaagsrersttsgqfgstts 1500

Db   1501 asnnrptflngvgaganltaalgvahssthegkpvgifpaftstnvsaalaldnrtsqsi 1560

Db   1561 slelkraepvtsndiseltstlgkhfkdsattkmlaalkelddakpaeqlhilqqhfsak 1620

Db   1621 dvvgderyeavrnlklklvirqqaadshsmelgsashsttynnlsrinndgivellhkhfd 1680

Db   1681 aalpassakrlgemmnndpalkdiikqlqstpfssasvsmelkdglreqtekaildgkvg 1740

```
Db   121 eaaapdaarltrsggvkrrnmddmagrpmvkggsgedkvptgqkrhqlnnfggmrqtmls 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   181 KMAHPASANAGDRLQHSPPHI 201
             ||||||||||||||||||||
Db   181 kmahpasanagdrlqhspphi 201


RESULT   5
Y75097
ID     Y75097 standard; Protein; 2514 AA.
XX
AC     Y75097;
XX
DT     21-MAR-2000  (first entry)
XX
DE     Neisseria meningitidis ORF 564 protein sequence SEQ ID NO:1668.
XX
KW     Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW     antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW     antibacterial; gene therapy.
XX
OS     Neisseria meningitidis.
XX
PN     WO9957280-A2.
XX
PD     11-NOV-1999.
XX
PF     30-APR-1999;  99WO-US09346.
XX
PR     01-MAY-1998;  98US-0083758.
PR     31-JUL-1998;  98US-0094869.
PR     02-SEP-1998;  98US-0098994.
PR     02-SEP-1998;  98US-0099062.
PR     09-OCT-1998;  98US-0103749.
PR     09-OCT-1998;  98US-0103794.
PR     09-OCT-1998;  98US-0103796.
PR     25-FEB-1999;  99US-0121528.
XX
PA     (CHIR ) CHIRON CORP.
PA     (GENO-) INST GENOMIC RES.
XX
PI     Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI     Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI     Tettelin H,  Venter JC;
XX
DR     WPI; 2000-062150/05.
DR     N-PSDB; Z53859.
XX
PT     Novel Neisserial polypeptides predicted to be useful antigens for
PT     vaccines and diagnostics -
XX
PS     Claim 2; Page 854-855; 1453pp; English.
XX
CC     Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent
CC     novel Neisseria meningitis and N. gonorrheae polynucleotides and
CC     polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR
CC     primers used in the exemplification of the present invention. The
CC     polypeptides, the polynucleotides, antibodies and compositions of
CC     the invention can be used as vaccines, as diagnostic reagents, and as
CC     immunogenic compositions. The polypeptides can be used in the
CC     manufacture of medicaments for treating or preventing infection due to
CC     Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC     presence of Neisseria bacteria, or to raise antibodies. They may also
CC     be used to screen for agonists or antagonists, which may themselves
CC     have use as antibacterial agents. The polynucleotides of the invention
CC     may also be used in gene therapy protocols.
XX
SQ     Sequence   2514 AA;

Query Match             2.9%;   Score 275;  DB 21;  Length 2514;
Best Local Similarity   18.8%;  Pred. No. 3.6e-08;
```

```
Matches  389;  Conservative  287;  Mismatches  803;  Indels  586;  Gaps  91;

Qy     7 GTEH-KAAVHTAAHNPV---------------------GHGVALQQGSSS-----------------SSPQNAAAS 43
          :||    |||  || |                            : ||      :||      ||:
Db    37 gsahvksvpfgtthapvcrsnifsfsllgfsiclavgtaniafadgiiadkaapktqgat 96

Qy    44 LAAEGKNRGKMPRIHQPSTAADGISAAHQQKKSFSLRGCLGTKFSRSAPQGQPG----- 98
          :|:  |        ||| |: :::  :::|  : :  |||  : :|   ||| ||||
Db    97 ilqtgng--ipqvnigtptsagvsvnqyaqfdvgnrgailnn--srsntqtqiggwigg 151

Qy    99 ------------------TTHSK-------------GATLRDLLARDDGETQHEAAAPDAARLTR 132
                                :||           : ||:|       :  |: | |
Db   152 npwlargearvvvnqinsshssqmngylevggrraevvianpaglavngggfinasratl 211

Qy   133 SGGVKRRNMDDMAGRPMVKG---GSGED-------KVPTQQKRHQLNNFGQMRQTMLS 180
          |  ||| |      :                      :  |  : :  |  |
Db   212 ttgqpqyqagdlsgfkirqgnvviaghgldardtfrilsyhskidapvwggdvrvvag 271

Qy   181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240
          : | |                         :           :
Db   272 qndvvatgna---hsp-----------------lnnaaantsnntanngthiplfaidty----- 312

Qy   241 LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300
                                            :|
Db   313 -----------------klg---gmyankitlista-------eqagirnqgqlfas 342

Qy   301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTOHYLAHHAS--SDGSQHLLLDN 358
          ||   |     :       :  |                  |       :|:
Db   343 sgn------vaidangrlvnsgtmaaan-----akdtdntaehkvnirsgg-----ven 385

Qy   359 KG----------HLFDIKSTATSYS---VLHNSHPGEIKGKLAQAGTGSVSVDGKSGKI 404
             ||          || |  |  :  :|||||
Db   386 sgtavsqggtqihsqsiqntgtllssgeilihns-------gslknetsgti------eaarl 435

Qy   405 SLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVW 464
          : | |            |  |         |:  :||
Db   436 aidtdt-lnnqgklsqtg----------------sqklhi------------------ 458

Qy   465 QSADKDTHSQLSRQADGKLYALKDNRTLQNLSD--------NKSSEKLVDKIKSYSVDQRGQ 518
          |                    |   |       |        | :         :  :
Db   459 -dagqkm----dnrgrmglqdtaptasngssnqtgn--synasfhss 498

Qy   519 VAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLV 578
          : |:|              |:        | |  |        |:|       :|  :  ||
Db   499 tttpttatgtgtatvsisnitapt------fad--gtirthgaldnsgsiiangqtd 547

Qy   579 VADSEGKLFSAAIPKQGDGNELKMAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN- 637
          |:  :              :      :  :        : ||  : | | | |:
Db   548 vsaqg--lnnag--qidihqlnakg--safdnhng--tiisdavhiqagslnnqngni 598

Qy   638 NFRQQHACPLGNDHQFHPGWNLTDALVID-NQLGLHHTNPEPHELLDMGHLGSLALQEGK 696
          |         |   |  |                    : | |  |     |  : |:  |
Db   599 ttrqq-----------------leietdqldnahgkllsaeiadlavsgslnnqnge 638

Qy   697 LHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLP 756
          | |     |  |           |  ||    : ||
Db   639 iatnqql-----------------------lihdgqqstavidntngtiqsgrdvaiqak 675

Qy   757 HVRNKPEPGDALQGLNK---------DDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLE 810
          : : |    |  | :|:               | |     :| :  | :| | ||
Db   676 slsn--ngtlaadnkldialqddfyvernivagne-lslstrgslknshtlqagkrir 730

Qy   811 RPAQTLSREG----ISGELKDIHVDHKQNLYALTHRGEVFHQPREAWQNGAESSWHKLA 866
          | | |  |:|       |  |: |   |          :   | |
Db   731 ikannldnaagqniqsggttdigtqhn------ltnrglidgq------ 767

Qy   867 LPQSESKLKSLDMSH------EHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTS 918
                                     : |   :||:|    :| | |||     |  |:||
Db   768 ---qtkiqagqmnnigtgriygdniaiatrldngdengtga--aiaarenlngig--- 819

Qy   919 GSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGMTGAEGRKVSSKFSERIRAYAFNPTMS 978
                |    |  |:  | |        |    ||   |:      :||
Db   820 ------qlnnrensliysgndmavggaldtnggatqkaqr 853
```

```
XX
AC   Y75098;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 564 protein sequence SEQ ID NO:1670.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic: diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2 .
XX
PD   11-NOV-1999 .
XX
PF   30-APR-1999;   99WO-US09346 .
XX
PR   01-MAY-1998;   98US-0083758 .
PR   31-JUL-1998;   98US-0094869 .
PR   02-SEP-1998;   98US-0098994 .
PR   02-SEP-1998;   98US-0099062 .
PR   09-OCT-1998;   98US-0103749 .
PR   09-OCT-1998;   98US-0103794 .
PR   09-OCT-1998;   98US-0103796 .
PR   25-FEB-1999;   99US-0121528 .
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,   Galeotti C,   Grandi G,   Hickey E,   Masignani V,   Mora M;
PI   Petersen J,   Pizza M,   Rappuoli R,   Ratti G,   Scalato E,   Scarselli M;
PI   Tettelin H,   Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; Z53860.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2: Page 861-862; 1453pp; English.
XX
CC   253015 to 254536, 254577 to 254615, and Y74253 to Y75941 represent
CC   novel Neisseria meningitis and N. gonorrheae polynucleotides and
CC   polypeptides. 254537 to 254576 and 254616 to 254673 represent PCR
CC   primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   2599 AA;
```

```
Query Match             2.6%;    Score 245;   DB 21;   Length 2599;
Best Local Similarity  19.0%;    Pred. No. 3e-06;
Matches  414;  Conservative  282;  Mismatches  766;  Indels  712;  Gaps  105;

Qy      10 HKAAVHTAAHNPVCHGVALQQGSSSSSPQNAAASLAAEGKNRGKMPRIHQPSTAAD--G   66
              |||||  |||  | | |  |    | ||    ||    :     |
Db     573 hdlavntqtaknsgh-lltqtgkidnrelhnagelaa----nnitlihsgrlsndkkgn 626

Qy      67 ISAAHQQKKSFSLR-------GCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGE 117
            | || |       |   | |||         |  |   |
Db     627 iraahlqldtaglhnagniladsgtvttknnlrntgkvsvarlntegqtldntrgrieae 686

Qy     118 TQHEAAAPDAARLTRSGG------------VKRRNMDDMAGRPM--------VKGG---SGE 156
```

```
RESULT  6
Y75098  standard;  Protein; 2599 AA.
ID  Y75098  standard;  Protein; 2599 AA.
```

```
Qy     979 TPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQL-----------------DAHNVRHNAP 1024
                      || |  :|: ||           |
Db     854 ----ihnagatieaagkmrlgveklhntnehlktqlvetgrehivdyeafgrhellregt 909

Qy    1025 QPDL------QSKLETLDLGEH----------------------GAELLNDMK 1049
            |  |      |    |   ||                      ||||  |
Db     910 qhelgwsvyndesdhlrtpdgaahenwhkydyekvtqktqvtqtapakiisgndlitidgk 969

Qy    1050 R-FRDELEQSATRSVTVLGQHQG----------VLKSNGEINSEFKPS-PGKALVQSFN 1096
            | | |  |      :  |  ||          | |  ||   |  | |     |  |
Db     970 evfntdsqiiagnlivqtekdglhneqtfgekkvfsengklhsywrekhkgr---dstg 1026

Qy    1097 VNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKT 1156
            :    | : |   |   |    |    ||  | :  || || |  ||   :|   :
Db    1027 hseqnytlpeeitrnislgsfayeshrkalshhapsqgtelpqsng---islpytsnsft 1083

Qy    1157 ALTKSRI-ILDTVTIGELHELADKAK-----LVSDHKPDA----DQIKQLRQQFDTLREKR 1207
            |  |:|  ||||     | | | ||      | | ||||     | | |  ::|   |
Db    1084 plpssslylinpvnkgylvetdprfanyrqwlgsdymldslkldpnnlhkrigdgyyeqr 1143

Qy    1208 YESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKK 1267
             :|       :     ||   | |    ||  |    :|    | |:   | ||
Db    1144 lineqiaelt--ghrrldgyyqndeeqfkalmdngataarsmnlsvgial---saeqvaq 1197

Qy    1268 LKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTS 1327
                       | | :                | || :  |||||  | | | | :: |
Db    1198 ltsdivwlvqke----vklpdgtqtvlvpqyvrvkngdidqkgal------lsgsntq 1247

Qy    1328 GGLNVSFGRDGGVSGNIMVATGHDVMPYMTGK---KTSA----GNASDWLSAKHKIS 1377
             ::     ||       |    | ||:: |    | |          |  | | ||
Db    1248 invsgslknsgtiagrnaliintdtldiniggrihaqksavtatqdinniggmlsaeqtil 1307

Qy    1378 PDLRIGAAV-SGTLQGTLQNSLKFKLTEDELPG-FIHGLTHGTLT----PAELLQKGIE 1430
            |  || |  ||  : |:|    |  |  |  |  | || |:             | | |
Db    1308 --lnagnninsqsttassqntqggsstyldrmagiyitgkekgvlaaqagkdiniiaggis 1365

Qy    1431 HQMKQGSKLTFSVDTSANLDL------RAGINLNEDGSKPNGVTARVSAGLSASA------ 1479
            :  |        | |  |            | :  |    |        | | ||
Db    1366 nqseegg-qtrlqagrdinldtvqtskhqathfdadnhvirgstnevgssiqtkgdvtlls 1424

Qy    1480 --NLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGI 1537
                  |  | |  : |:       ||   ||     |  |:|  | |     |   | |
Db    1425 gnnlnakaaevssangtl--avsaknd----in-isagintt--hvddaskhtgrsgg- 1473

Qy    1538 FPAFTSTNVSAAALALDNRTSQSISLELKRAEPVTSNDISELTSTL----------GKHFK 1587
            |  | |          |  |: |    :     |::  |   | |           || |
Db    1474 --gnklvitdkaqshhetaqsstfegkqvvlqagndanilgsnvisdngtqigaqnhvr 1530

Qy    1588 DSATTKMLALALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSH 1647
              |  |  |  | | |  |   |     |     || |      || |    : :
Db    1531 igtt--------qtgsqsetyhgtqksglmsaig----------fti--gsktn 1565

Qy    1648 SMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMM----NNDPALK 1702
            |    |:  | | | :        |::  | |       || | :  |    |:|  ||
Db    1566 tqenqsqsnehtgstvgslkgdtti-vagkhye------qigstvspegnntiyaq 1615

Qy    1703 DIIKQLQSTPFSSASVSMELKDGLREQTEKAIILDGKVGREEVGV-----LFQDRNNLRV 1756
            | | |     |   |   :| | |    ||   |    | |  |     |  |  | |  |
Db    1616 sidiqaahklnsnttqtyeqkgltvafsspvtd--laqqaiavaqsskqvgqsknd-rv 1672

Qy    1757 KSVSVSQS--VSKSEGFNTPALLGTSNSAAMSMERNIGTINFKYGQDQNTPRRFT-LEG 1813
              |  | |    |  | :||     |||   |  || | ||||| |  |:||||  ||
Db    1673 namaaaanagwqavqtgksaqnlangttnakqvs-----isitygeqqn--rqttqvga 1723

Qy    1814 GIAQANPQVASALTDLKKEGLEMKS 1838
            | | |   | | :|    | :||
Db    1724 nqaqasqiqagqkttllatgaaeqs 1748
```

```
Db   687 |:------||||:|||:  |:  ||                                        742
         tvn----igsqqltnqsghitaeqltinsrvdnqngkllsanqaqlavsdglynqhge

Db   157 DKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGST   216

Qy   743 iatnrqlisihdkn-----qntlalnnadgtiqsagn-----------vslqaksla    782

Db   217 SKATTAHADRVEIAQEDDSEFQQLHQQRLARERENPPQQPKLGVATPISARFQPKLTAV    276

Qy   783 nngtltagnkldialtdd--------fvverd-------------ltagkqlnls-     816

Qy   277 AESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGK   336

Db   817 ikqrlknthtlq------aght--lklnagni--dn--------qvtgk             847

Qy   337 DTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVS   396

Db   848 -----iiggeqtditseqhvdnrg--------lins-------dgithigagqtl       882

Qy   397 VDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPGGE--SIRLHDDKI   454

Db   883 tntgtgki-ygnhialdaqillnreettegstkag----aiaarkrldigakeihnqeg   936

Qy   455 HILHPE-LGVWQSADKDTHSQ----------LSRQADGKLYALKDNRTLQNLSDNKS     500

Db   937 allssegifavgnrldeqhhaagmadtfvngsaglevqgd---almsvrnmqninnhfk   992

Qy   501 SE---KLVDKIKSYSV----DQRGQVAILTDTPGRHKMSIMPSLDASPESHISLS-L     549

Db   993 tetylakaekqvrdytvlgqntyyqagkdglfdnsqq-kdqttathlkngsrieanqw  1051

Qy   550 HFADAH-----QGLLH----------GKSELEAQSVAISHGRLVVAD-------       581

Db  1052 hvrdyhietykerilenrpahitvggdltassgqnwlnkdsrivvggrlitddlnqkeitn 1111

Qy   582 --SEGKLFSAAIPKQGD--------GNELKKMAMPQHALDEHFGHDHQISGFFHDDHGQ   630

Db  1112 qsttgkgrtdavgtqwdsvtkkgwysgrkrqrrternhtp---yhdtql--fthdfdtp  1165

Db   631 LNALVKN---NFRQQHAC------PLGNDHQFHPG--WNLTDALVIDNQLGLHHTN      675

Qy  1166 vsviqqnaaspsfqpaasaiklidgystaavnggrihtgnvvslnnatvtlpnsslytth 1225

Qy   676 PEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKR  735

Db  1226 pd----nkgw-----lvetdpqfady--rrwlgsdymlqqlqldtnhlhkrlgdgyyeq 1273

Qy   736 LNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKG  795

Db  1274 klvn--------eqihqltgyrl-------ldgyrsdee------qfkalmdng      1306

Qy   796 --DIRSFQIKPG----TQQ--------LERPAQTLSREGISGELKDIHVDHKQNLYAL    839

Db  1307 ltaaktfgltpgialsaeqvarltsdivwmenqtvtls---dgstqtvlv--pkvyal  1359

Qy   840 THEGEV------FHQPREAWQNGAESSS------WHKLALPQSES------         872

Db  1360 arkgdintsgglisaeqvllklqngnltnsgtiagrqavliqarninsngniqadqiglk 1419

Db   873 KLKSLDMSHEHKPIATFEDGSQHQ---------LKAGGWHAYAAPER------G       911

Qy  1420 aeksini---------dggqvqagrlltaqaqninlngttqtsgnerngntaidrmag  1468

Qy   912 PLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGMTGAEGRKVSSKFSE--RI   968

Db  1469 invvg-shteqvdnrtsdji-------lslhasndininaatvsnqvkdgttqi      1514

Qy   969 RAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDL 1028

Db  1515 ta-gnnlngtlirtehreaygt---------lddenhrhvrqstev               1550

Qy  1029 QSKLETLDLGEHGAELL--NDMKRFRDELEQ------SATRSVT------VL        1066
```

```
Db  1551 gssirt---qngallragndlkirggeleaeegktvlaagrdvtisegrqiteldtsvs 1606

Qy  1067 GQHQGVLKSNGEINSEFKPS-----------PGKALV---QSFNVN-            1101

Db  1607 gkskgilsst-kthdryrfshdeavgsniggqkmivaagqdinvrgsnlisdkgivlkag 1665

Qy  1102 QDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQR--DPNDKTALT 1159

Db  1666 hdidistahnrytgneyheskksqvmg---tgglgft------ignrkttddtdrtniv 1715

Qy  1160 KSRLIL-----DTVTIGELHELADKAKLVSDHKPDADQIKQLRQQPFDTLREKRYESNPVK 1214

Db  1716 htgsilgslngdtvtvagnryrqtgstvsspegrntvtaksidvef---annryatdyah 1772

Qy  1215 HYTDMGFT----HNKALEANYDAVKAFINAFKKEHHGVNLTRTVLESQGSAELAKKLKNT 1271

Db  1773 tqeqkgltvalnvpvqaagnfiqaaqnvgksnkrvnamaaanaawq-syqatqqmqqf  1831

Qy  1272 LLSLDSGESMSFSRSYGGGVSTVFVPTLSK-------KVPVPVIPGAGITLDRAYNLSF  1323

Db  1832 apssagqgqnnqspsisvsityyeqksrneqkrhyteaaasqiiqkgqt------tl   1884

Qy  1324 SRTSGG----LNVSFGRDGGVSGNIMVATGHDVM--PYMTGKKTSAGNASDW-LSAKHKI 1376

Db  1885 aatgsgeqsninitgsdvighagtaliadnhirlqsakqdgseqsknkssgwnagvavki 1944

Qy  1377 SPDLRIGAAVSGTL-QGTLQNSLKFKLTEDELPGFIHGLTH-----------GTLT    1420

Db  1945 gngirfgitaggnigkgkegg-------------sthrhthvgsttgkttirsggdttlk 1993

Qy  1421 PAELLQKGIE---------------------HQMKQ------GSKLT----        1440

Db  1994 gvqligkglqadtrnlhiesvqdtetyqskqngnvqvtvgyfsasgsyrqskvkadha  2053

Qy  1441 -------------FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSR  1486

Db  2054 svtgqsglyagedgqyqikvrdntdlkggiltssqsaedkgknlfqtatltas-diqnhsr 2112

Qy  1487 ERSTTSGQFGS----------TTSASNNRPTFLNGVGAG-------ANLTAALGV---AH 1526

Db  2113 yeqrsfgiggsfdlnggwdgtvtdkgrptdrispaaqyqsgdgsknsttrsgvnthnih 2172

Qy  1527 SSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSISLELKRA-EPVTSNDIS---ELTST 1581

Db  2173 itdeaqqlartgrtaketeariytgidtetadqshsghlknsfdkdavakeinlqrevtke 2232

Qy  1582 LGKHFPDSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQ 1641

Db  2233 fgrn------aaqavaav-----eryqeartl-leael                      2268

Qy  1642 QAADSHME-----------adklgntqsy------LGSASHSTTYNNLSRINN       1668

Db  2269 qntdseaekaafraslgqvnaylaenqsrydtwkeggigrsilhgaaggilttgslggila 2328

Qy  1669 DGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLRE 1728

Db  2329 gggtslaapyldka---aenlg------pagkaavnalggaaigyat---------    2366

Qy  1729 QTEKAILDGKVGREEVGVLFQDRNN--LRVKSVSVS-----QSVSKSEGFNTPALLL-  1778

Db  2367 ---ggsgavvganv-dwnnrqlhpkemaladkyaealktevekregrki----       2412

Qy  1779 GTSNSAAMSMERNI 1792

Db  2413 -ssqeaamrirrqi 2425
```

```
DT   07-JUL-2000 (first entry)
XX
DE   Human homologue of UNC-53 (Hs-UNC-53/1) sequence.
XX
KW   UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
KW   anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
KW   antisclerotic; antimetastatic; anti-arthritic; autoimmune disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          958..1014
FT                   /note= "this region is found to be absent when encoded by
FT                   a variant cDNA isolated from frontal cortex"
FT   Region          1033..1040
FT                   /note= "this region is found to be absent when encoded by
FT                   a variant cDNA isolated from frontal cortex"
FT   Region          1173..1175
FT                   /note= "this region is found to be absent when encoded by
FT                   a variant cDNA from Hela or colorectal
FT                   adenocarcinoma tissue"
FT   Misc-difference 1233
FT                   /label= Leu or Ser
XX
PN   WO9963080-A1.
XX
PD   09-DEC-1999.
XX
PF   02-JUN-1999; 99WO-EP03848.
XX
PR   03-JUN-1998; 98GB-0011962.
XX
PA   (JANC ) JANSSEN PHARM NV.
XX
PI   Luyten WHML,  De Raeymaeker MC, Geysen JJGH,  Bogaert TAOE;
PI   Maerten LJS,  Verhasselt P,  Van De Craen M;
XX
DR   WPI; 2000-116370/10.
DR   N-PSDB; A07835.
XX
PT   Novel proteins and nucleic acids e.g. for treating neurodegeneration -
XX
PS   Claim 93; Fig 1b; 146pp; English.
XX
CC   The invention provides vertebrate (human) protein homologue of a UNC-53
CC   protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
CC   their plus ends. The UNC-53 sequences are used to promote neural
CC   regeneration, revascularization and wound healing; also for treating
CC   neurodegenerative disease, acute traumatic injury, fibrotic disease and
CC   autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
CC   polynucleotides can be used for recombinant production of the proteins,
CC   as a source of probes for detecting allelic variants and polymorphisms,
CC   for sequencing genomic DNA and for detecting UNC-53 expression; and as
CC   source of therapeutic antisense sequences. Cells that express the
CC   protein are used to identify regulators of cell shape, growth, motility
CC   and migration. They can also be used to identify proteins that are
CC   involved in signal transduction pathways also involving UNC-53, and to
CC   identify compounds that alter attachment of UNC-53 to microtubules. A
CC   target gene coupled to a UNC-53 encoding sequence may be used to deliver
CC   the target gene to a cellular microtubule or its plus ends. The present
CC   sequence represents the amino acid sequence of the first human homologue
CC   of UNC-53, designated hs-UNC-53/1.
XX
SQ   Sequence 1837 AA;

Query Match            2.6%;  Score 241.5; DB 21; Length 1837;
Best Local Similarity 20.3%;  Pred. No. 2.9e-06;
Matches 345;  Conservative 221;  Mismatches 677;  Indels 459;  Gaps 78;

Qy     93 PQGQPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLTRSGGVKRRNMDDMAGRP--MV 150
          :|:  |||  |  ||   |   |||  |  |     ||  :|:|    |  ||
Db     36 PggppasnlrkqkstlnlsfltdsekklqIyepews-----------ddmakapkglg  82
```

```
Qy   151 KGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAH-------------------PASANAGD 192
         ||| | |||                                            ::
Db    83 kvgskgreaplmsk-----------------------tlsksehslfgakgspaggaktplaplaplg- 128

Qy   193 RLQHSPPHIP-GSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQLHQQRLARERE 251
         : ::|    | |                                   ||
Db   129 ---kpsriprgpyaevk-----plskap--------eaavsedgksddellsskakqks 172

Qy   252 NPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAVTL 311
         : |||          |       |   |   |    |      |    | | :|
Db   173 sgpvpsakqqeeraflkvdpelvvtvlgdleqllfsqm-ldpesqrktvqnv----l 225

Qy   312 DKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQ--HLLLDNKGHLFDIKSTA 369
         |                       |  : | :|    |      |   :|| |  |
Db   226 d---------lrqnleetm-----slrgsqvthsslemtcydsddanpr 261

Qy   370 TSYSVLHNSHP----GEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQPGE- 423
         :|:  |   |      |:|   |   |   | |    |        ::    : ||
Db   262 svsslsnrsspIswryggsprI-qagdap------svggscrsegtpawymhger 310

Qy   424 AHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADK-DTHSQLSRQAD- 480
         ||                                  :       |   | |     |
Db   311 ah-------yshtmpmrspskls------hisrlelveslddevdlksgymsdsdlm 355

Qy   481 GKLYALKDNRT--------------LQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDPGR 528
         ||:  |  |                |::   |        |:  :|     |   | |
Db   356 gktmteddiittgwdesssissglsdasdnlsseefnasslnslps-------tptasr 408

Qy   529 HKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHG----RLVVAD 581
         |    |  ||    |    |  |   |  | |  |  :|    |         |:| |
Db   409 rnstivlrtdsekrslaesglswfseseeakpkleydsgsIkmepgtskwrrerpescd 468

Qy   582 SEGKLFSAAIP-KQGDGNELKMKAMPQHALDEHFGHDHQ---ISGFFH--DDHGQLNA 633
         |  |||       |  |  | | |   |     | ||       |        | |
Db   469 dsskggelkkpislghpgslkkgktppvavtspithtaqsalkvagkpeqkatdkgkl-- 526

Qy   634 LVKNNFRRQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTN----PEPHEILDMGH 686
         |  |          | |  | | |   |  |::     | |       |  |  | |
Db   527 avkntglqrsssdagrd-----rlsdakppsgiarpstsgsfgykkppp--------a 572

Qy   687 LGSLALQGEGKLHYFDQLTKGWTGAESDCKQLKK-GLDGAAYLLKDGEVKRLNINQSTSS 744
         ||                                    |||   |  |  |        :
Db   573 tgtatvmq-----------------tggsatlskiqkssgi---------pvkpvngrktsld 609

Qy   745 IKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIG-------VNKYLALTE 793
         :  |   |  |  |                 |  |              |   | |
Db   610 vsnsaepgflapgarsniq---yrslprpaksssmsvtggrggprpvsssidpsllstk 665

Qy   794 KGDIRSFQIKPGT-----QQLERPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFHQ 848
         |   |  | |      |   |     | |     :  |    |        |   | | |
Db   666 ggqltpsrlkeptkvasgrttpapvnqtdrekekakakavald-sdni-slksig---s 719

Qy   849 PREAWQNGAESSWHKLA-LPQSESK----------LKSLDMSHEHKPIATFEDGSQHQ 896
         |     |    | |                     |                      |
Db   720 pestpknqashptatklaelpptplrataksfvkppslanldkvnsnsldlpsssdttha 779

Qy   897 LKAGGWHAYAAPERGPL-AVGTSGSQTVFN---RLMQGVKGKVIPGSGLTVKLSAQTGG 951
         |  |          | |  | ||   | |     | ||       |    |       ||
Db   780 skvpdlhatssasgglpscftpspapilninsasfsgglelm---sgfsvpketrmyp 835

Qy   952 MTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQH----GWQG--REGLKPLY 1004
         |   | |  |       |   |          |        |      |         |
Db   836 klsglhrsmeslqmpmslpsafpsstpvptppappaapteeeteeltwsgsprag---- 890

Qy  1005 EMQGALIKQLDAH-NVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSV 1063
         |                |    | |     |       |    |:     : |
Db   891 ------qldsnqrdrntlpkkglryqlgsqe--------etkerrhs 923

Qy  1064 TVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATP--MV 1116
         ||  | |  ||    || |      ||     :      |  |  | : ||
Db   924 htig---glpesddq--sel-psp-palpms--lsakqq-ltnivsptaattpritrsns 973
```

```
N-PSDB; V03553.
DR
XX
PT  Genes present in Neisseria meningitidis but not other Neisseria
PT  species - and related host cells, RNA, anti-sense sequences,
PT  polypeptide(s) and antibodies, useful for diagnosing Neisseria
PT  meningitidis infection and in protective vaccines
XX
PS  Claim 8; Pages 81-90; 150pp; French.
XX
CC  W42633-37 and W42639-41 are encoded by a DNA sequence found in region 2
CC  of Neisseria meningitidis. The specification describes DNA sequences
CC  that are found in N. meningitidis, but not in N. gonorrheae or
CC  N. lactamica, except for the genes involved in biosynthesis of the
CC  capsule polysaccharide, frpA or C, opc, porA, rotamase, sequence IC1106,
CC  IgA protease, pillin, pilC, proteins which bind transferrin and opacity
CC  proteins. The DNA sequences are found on chromosome 22491, mainly (or
CC  within 20 kb) between tufA and pilT (region 1), pilQ and lambda-740
CC  (region 2) or argF and opaB (region 3). The DNA sequences are responsible
CC  for the differences in pathogenicity between N. meningitidis and
CC  N. gonorrheae, specifically they include the genes that allow
CC  N. meningitidis to cross the blood-brain barrier. DNA sequences common
CC  to N. meningitidis and N. gonorrheae, but absent from N. lactamica, are
CC  responsible for colonisation and penetration of the mucosa. The DNA
CC  sequences can be used to produce probes and primers, and antibodies
CC  produced against the encoded proteins are used in standard
CC  hybridisation/immunoassay processes for diagnosis of N. meningitidis
CC  infection, particularly meningitis.
XX
SQ  Sequence  1981 AA;


Query Match          2.5%;   Score 232;  DB 19;   Length 1981;
Best Local Similarity  19.4%;   Pred. No. 1.3e-05;
Matches 380;  Conservative 271;  Mismatches 712;  Indels 600;  Gaps 92;

Qy  309  VTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLDNKGHLFDIKST 368
Db   75  ittdks---apknqqv--ilktntgap---lvniqtpngrg--lshrytqfdvdnk 122

Qy  369  ATSYSVLHNSHPGEIKGK--------------LAQAGTGVSVSVDGKSGKISL------GSG 409
Db  123  gavlnndrnnnpflvkgsaqlilnevrgtcasklngivtvggkadvilanpngitvnggg 182

Qy  410  TQSHNKTMLS--QPGEAHRSLLTGI----------WQHPAGA---- 441
Db       ------------AR 441
Db  183  fknvgrgiltigapqigkdgaltgfdvrqgtltvgaagwndkggadytgvlaravalggk 242

Qy  442  PQGESIRLH--DDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTL----- 492
Db  243  lggknlavstgpqkvdyasgeisagtaagtkptialdtaalggmya--dsitlianekgv 300

Qy  493  --QNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLH 550
Db  301  gvknagtleaakgli-vtssgriensgriattad----------gteasp-tylsie- 345

Qy  551  FADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHAL 610
Db  346  --------ttekgaagtfisnggri-eskgll----vietgedislrngavvqnng 388

Qy  611  DE-------HFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLG-----NDHQFHPGWNL 659
Db  389  srpattvlnaghnlvi------esktnv---nnakgsanlsaggrttindatiqagssv 438

Qy  660  TDALVIDNQLGLHHTN--PEPHEIDMGHLGSLALQEGK----------696
Db  439  ysstkgdtelg-entrilaenvtvlsngsigsaavieakdtahiesgkpisletstvasn 497

Qy  697  ----------LHYFDQLTKGWTGAESDCK-------------QLKKGLDGAAYLLKD 730
Db  498  irlnngnikggkqlalladdnitakttnltnpgnlyvhtgkdlnlnvdkdlsaasihlks 557

Qy  731  ------GEVKRL----------------NINQSTSS----IKHGTENVFSLPHVRN--- 760
```

```
Qy 1117  -PSAESKLQSMLGHFVSAGVDMSHQ-KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELH 1174
Db  974  iptheaafelysgsqmgstislaerpkgmirsgsfrdptddvh--------gsvl 1020

Qy 1175  ELADKAKLV--SDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANY 1231
Db 1021  slassasstyssaeermqseqirklrrelessqek----------vatltsqlsana 1067

Qy 1232  DAVKAFINAFKKEHHGVNLTRTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYGGGV 1291
Db 1068  nlvaaf------eqslvnmtsrl---rhlaetaee-kdtel-ldiretidflkkknsea 1115

Qy 1292  STVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHD 1351
Db 1116  qaviqgalnasettp---------kerlrlkrqnssdsis--slnsitshssigsskd 1162

Qy 1352  VMPYMTGKKT-----------SAGNASD----WLSAKHKISPDLRIGAAV 1386
Db 1163  adakkkkkswvyelrssfnkafsikkgpksassysdieeiatpdssapsspklqhgs- 1220

Qy 1387  SGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPA-------ELLQKGIEHQMK 1434
Db 1221  tetaspsiksstxssvgtdvtegpahpaphtrlfhaneeeepekkevselrselwekemk 1280

Qy 1435  QGSKLTFSVDTSANLD-LRA---GINLNEDGSKPNGVTARVSAGLSASANLAAGSRERST 1490
Db 1281  ltdirlealnsahqldqlretmhnmqlevdllkaendrlkvapgpss--------gs 1329

Qy 1491  TSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAAL 1550
Db 1330  tpgqvpgssalssprrsl-----glalthsfgpsladtdlspmdgistcgpkeevtlrv 1383

Qy 1551  ALDNRTSQSISLELKRAE---------------PVTSNDISEL--TSTLGKHF 1586
Db 1384  vvrmppqhiikgdlkqqefflgcskvsgkvdwkmldeavfqvfkdyiskmdpastlgis- 1442

Qy 1587  KDSATTKMLAALKELDDAKPAE 1608
Db 1443  tesihgysishvkrvldaeppe 1464

RESULT  8
W42634
ID   W42634 standard; Protein; 1981 AA.
XX
AC   W42634;
XX
DT   22-OCT-1998 (first entry)
XX
DE   Protein sequence that is specific for Neisseria meningitidis.
XX
KW   N. gonorrheae; N. lactamica; chromosome 22491; region 1; region 2;
KW   region 3; pathogenicity; blood-brain barrier; diagnosis; infection;
KW   meningitis.
XX
OS   Neisseria meningitidis.
XX
PN   WO9802547-A2.
XX
PD   22-JAN-1998.
XX
PF   11-JUL-1997; 97WO-FR01295.
XX
PR   12-JUL-1996; 96FR-0008768.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA   (SMIK ) SMITHKLINE BEECHAM.
XX
PI   Nassif X, Tinsley C, Achtman M, Merker P, Ruelle J;
PI   Vinals C;
XX
DR   WPI; 1998-110594/10.
```

```
Db   558  dnaahitgsktltaskdngveagllnvtntnlrtnsgnlhiqaakgni---qlrntkl 613
          ||:|  |  |  :    |     | :   :       :|      :      :|||:
Qy   761  --KPEPGDALQGLNKDDKAQAMAVIGVNKYLA------LTEKGDIRSFQIKPGT 806

Db   614  naakalettalqgnivsdglhavsadghvsllangnadftghntltakaдvnagsvgkgr 673
          |:      ||| |: |   | | :       |
Qy   807  QQLERPAQTLS-------------REGISG-------ELKDIHVDHKQ- 834

Db   674  lkadntnitsssgditlvagngiglgdgkqrnsingkhisiknggnadlkninvhaksg 733
          | |  | |    |  ||  | :    | |        | :   ||
Qy   835  --NLYA-------LTHEGEVFHQPREAWONGAE-----SSSWHHKLALPQ 869

Db   734  alnihsdralsientklesthnthinaqhervtlnqvdayahrhlsitgsqiwqndklp- 792
          :     |  :  : :       |  |  :          |   :|| |  :|
Qy   870  SESKLKS---LDMSHEHRPIATFEDGSQHLKAGGWHAYAAP--ERGPLAVGTSGSQTV 923

Db   793  sanklvangvlalnarysqia---dnttlragaintagtalvkrgninwstvstktl 847
          |     | | |:|       ::     || |      : |
Qy   924  -FNRLMQGVKGK--VIPGSG-LTV--KLSAQT-----------GGMTGAEGRKV 960

Db   848  ednaelkplagrlnieagsgtltiepanrisahtdlsiktggkillsakggnagapsaqv 907
          :           |   | |   || |         |  | |  ::  |:    ||  | |
Qy   961  SS---KFSERIRAYAFNPTMSTPRPIKNAAYATOHGWOQREGLKPLYEMQGALIKQLDAH 1017

Db   908  ssleakgnirlvtgetdlrgskitagknlvvattkgklnieavnsfsnyfptqkaael- 966
          ||         |  |          :  |      | ::    :    |:     :|
Qy  1018  NVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELE---QSATRSVTVLGQHQGVLK 1074

Db   967  ----nqkskeleqqiaqlkksspkskliptlqeerdrlafyiqainkevk--gk------k 1015
              :| ||      |   |::        ::|    |     :|||      |:        |
Qy  1075  SNGEINSEFKPSPCKA-LVQSFNVNRSGQDLSKSLQQAVHAT--PPSAESKLQSMLGHF 1130

Db  1016  pkgkeylqaklsaqnidlisaqgieisgsditaskklnlhaagylpkaadseaaail--- 1072
          |        |  |:       |   |  |  :  | |  :      |  :     ||
Qy  1131  VSAGVDMSHQKGEIPLGRQRDPNDKTALTK-SRLILDTVTIG-ELHELA--DKAKLVSDH 1186

Db  1073  ----idgitdqyeigkptykshydkaalnkpsrl--tgrtgvsihaaaaalddariil-- 1123
              |        |  |     |     ||         ||  |  |    | ||
Qy  1187  KPDADQIKQLRQQFDTLREKRYESNPVKHYHYTDMGFTHNKALEANYDAVKAFINAFKKEHH 1246

Db  1124  --gaselkapsgsid-------ikahsdi------vleaqqndaytflktkgksgk 1164
            | |     | |          ||  |        ||       | | ||| ||
Qy  1247  GVNLTTRT----------------VLESQGSAELAKKLKNT--LLSLDSGESMSF- 1283

Db  1165  iirktkftstrdhlimpapveltangitlgaggnieanttrfnapagkvtlvageelqll 1224
          :  :|          | ||  |    :||   ::|      :           ||  :|||
Qy  1284  --------------SRSYGG--GVSTVFVPTLSK-KVPVPVIPGAGITLDRAYNLSFSR 1325

Db  1225  aeegihkheldvqksrrfigikvgksnysknelnetklpvrvvaqtaatr--------- 1274
          : |||            | ||| |:|     : || :|   |    ||||
Qy  1326  TSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAG------NASDWLS 1371

Db  1275  --------sgwdtvlegteikttlagadiqagvgekarvdakiillkgivnriq 1319
                  | | :|   ||    || |  | | | ||     || ||  :|| ||
Qy  1372  AKHKISPDLRIGAAVSGTLQGTLQNSLKFKLITEDELP------GFIHGLTHGTL-TPAE 1423

Db  1320  seekletnstvwkqag--rgstietlklpsfesptppklsapggyivdipkgnlkteie 1377
          |:  |  |  |  :||     ||  |  || |    | |  :||| |  ||||  ||:|
Qy  1424  LLQKGIEH-QMKQGSKLTFSVDTSANLDLRAGINLNED------GSKPNGVTAR---- 1470

Db  1378  klskqpeyaylkq-----------lqvakninwnqvqlaydrwdykqegliteagaali 1424
          :|| |||  |||              || | :|:  | || |||||   | |||| ||
Qy  1471  ----VSAGLSASANLAAGSRERSTTSGOFGSTTSASNNRPTFLNGVCAGANLTAALGV 1524

Db  1425  alavtvtvsggtgavlglngaaaaatdaafasl--asqasvsfinnkgvgktlkelg- 1481
          |||||||| ||| ||  |::      || ||||    |||| ||:||||| |||| |
Qy  1525  AHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQ---SISLELKRAEPVTSNDISELTS 1580

Db  1482  -rsstvknlvvaatagvadkigas-alnnvsdkqwinnltvnlanagsaalin----ta 1535
           ||| |  |||| || |  |  |   |||| |||| ||||||| ||||||||||    ||
Qy  1581  TLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAKDV---- 1622

Db  1536  vnggslkdnleanilaalvntahgeaaskikqldqhyivhkiahaiagcaaaankgkcq 1595
```

```
Qy  1623  ------VGDERYEAVRNLKK----------LVIRQQAADSHSMELGSASHSTTYNNL 1663
          ||:|   |:     :| |     |:   :    : :   : :  :| :
Db  1596  dgaigaavgeivgealtngknpdtltakereqilayskivagtvsgvvggdvnaaanaae 1655

Qy  1664  SRINNDGIVELLKHFDAALPASSAKRLGEMNNDPAL--KDIIKQLQSTPFSSASVSME 1721
          : | | : :    | :||  :|   : | |:  :      ||::| | ||| : ::|||
Db  1656  vavknqlsdkegrefdnemtacakq-----nnpqlcrkntvkkyqnvadkrlaasia 1708

Qy  1722  LKDGLREQTE-----KAILDGKV-GREEVGVLFQDRNNLRVKSVSVSQS-VSKSEGF 1771
          ::||  | |      ||:: :   |||| |:| |  : | : |  :::  ||: |
Db  1709  ictdisrstecrtirkqhiidsrslhssweagligkddewylfsksytqadalqsyhl 1768

Qy  1772  NTPA-LLLGTSNSAAMS---------------MERNIGTINFKYGQ--- 1801
          ||   : |:|:||: :|               ||:| |||| ||||
Db  1769  ntaakswlqsgntkplsewmsdqgytlisgvnprfipiprgfvkqntpitnvkypegisf 1828

Qy  1802  DQNTPRR--------FTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
          | | :||        |:|| |: | |  | | |  ||:||| :|
Db  1829  dtnlkrhlanadgfsqeggikgahnr-tnfmaelnsrggrvks 1870
```

```
RESULT  9
Y85574
ID   Y85574 standard; Protein; 2608 AA.
XX
AC   Y85574;
XX
DT   07-JUL-2000 (first entry)
XX
DE   Hs-UNC-53/3/GFP fusion insert of plasmid pGI3303.
XX
KW   UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
KW   anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
KW   antisclerotic; antimetastatic; anti-arthritic; autoimmune disease.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1194
FT                   /note= "unspecified"
XX
PN   WO9963080-A1.
XX
PD   09-DEC-1999.
XX
PF   02-JUN-1999; 99WO-EP03848.
XX
PR   03-JUN-1998; 98GB-0011962.
XX
PA   (JANC ) JANSSEN PHARM NV.
XX
PI   Luyten WHML,  De Raeymaeker MC,  Geysen JJGH,     Bogaert TAOE;
PI   Maerten LJS,  Verhasselt P,  Van De Craen M;
XX
DR   WPI; 2000-116370/10.
DR   N-PSDB: A07846.
XX
PT   Novel proteins and nucleic acids e.g. for treating neurodegeneration
XX
PS   Disclosure; Fig 7e; 146pp; English.
XX
CC   The invention provides vertebrate (human) protein homologue of a UNC-53
CC   protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
CC   their plus ends. The UNC-53 sequences are used to promote neural
CC   regeneration, revascularization and wound healing; also for treating
CC   neurodegenerative disease, acute traumatic injury, fibrotic disease and
CC   autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
CC   polynucleotides can be used for recombinant production of the proteins,
CC   as a source of probes for detecting allelic variants and polymorphisms,
CC   for sequencing genomic DNA and for detecting UNC-53 expression; and as
CC   source of therapeutic antisense sequences. Cells that express the
```

```
CC   protein are used to identify regulators of cell shape, growth, motility
CC   and migration. They can also be used to identify proteins that are
CC   involved in signal transduction pathways also involving UNC-53, and to
CC   identify compounds that alter attachment of UNC-53 to microtubules. A
CC   target gene coupled to a UNC-53 encoding sequence may be used to deliver
CC   the target gene to a cellular microtubule or its plus ends. The present
CC   sequence represents the amino acid sequence of a full-length Hs-UNC-53/3
CC   in fusion with GFP insert of plasmid pGI3305.
XX
SQ   Sequence   2608 AA;

Query Match     2.4%;   Score 225;   DB 21;   Length 2608;
Best Local Similarity   18.7%;   Pred. No. 5.4e-05;
Matches 406;   Conservative 290;   Mismatches 727;   Indels 752;   Gaps 107;

Qy     22 VGHGVALQQGSSSSSPQNAAASLAAE-------GKNRGKMPRIHQP-------STA 63
             |||
Db    448 vthasppseasqaktqqdmqsslaaryatqsnhsgiatsqkkpterlpgpsrvpaagsssk 507

Qy     64 ADGISAAHQQKSFSL---RGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDG 116
             ::| |  ::|::|
Db    508 vggasnlrrsgsfnsidknkppnyangnekdsskgpqsssgv----ng 552

Qy    117 ETQHEAAA-----------PDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRH 166
             :|        ||  |:|| | |
Db    553 nvqppstagqppasaipspsaskpwrs---ksmnvkhsatstml---- 593

Qy    167 QLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADR 226
                            |:||
Db    594 ------tvkqsstatsptpss----drlk--ppvseg----vktapsgqks---------- 628

Qy    227 VEIAQEDDDSEFQQLHQQRL-ARERENPPQPPPKLGV------------------ 261
                                       |||||
Db    629 -------------mlekfklvnartalrppqppssgpsdggkdddafsesgemegfnsgl 675

Qy    262 ---ATPISARFQPKLT----------AVAESVLEGTD--------TTQSPLKPQ--S 295
              :|| :  ||||            :::| ||||         ||:::||  S
Db    676 nsggstnsspkvspklappakagsknlsnkksllqpkekeeknrdknkvctekpvkeekdq 735

Qy    296 MLKGSGAGVTPLAVTLDKG--------KLQLAPDN----PPALNTLLKQTL------GKD 337
                                       |:|:|||    ::|::|:  |         GKD
Db    736 vtemapkktskiaslipkgskttaakkeslipsssgipkpgskvptvkqtispgstaske 795

Qy    338 TQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLA---QAGTG 393
             :::|                         ||:|                    QAGTG
Db    796 ------------iskpitmekasasscpaplegreagqaspsgsc 842

Qy    394 SVSVDGKSGKISLGSGT-------QSHNKTMLSQP--GEAHRSLLTGIWQHPAGAARPQ 443
                                        ||::|||:||
Db    843 tmtvaqssgq-stgngavqlpqqqghspntatvapflyrah-------senegtalps 893

Qy    444 GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSD-NKSSE 502
             :::: |     |:::   :|::||:|::|:::||:||::  ::::|::|||:|:|  NKSSE
Db    894 adsc-tsptkmdlsysk--------takqcleeisgedpetrrmrtvkniadirqnle 942

Qy    503 KLVDKIK---SYS-VDQRGQVAILTDTPGRHKMSIMPSLLDASPESHISLSLHFADAHQG 557
             :|::|:|   |:|                             :|::||:|||:|:::|
Db    943 etmsslrgtqishstlettfdstvttevngr------tipnltsrp---tpmtwrlgqacpr 995

Qy    558 LLHGKSELEAQSVAISH---GRLVVADSEGKLFS-----AAIPKQGDGNELKMKAMPQ 607
             ::|                  ::|:|::||::||     :|:||||| :|:|:|:|:|Q
Db    996 lqag-----dapslgagyprsgtsrfihtdpsrfmyttplrraavsrlgnnmsqidmsekas 1051

Qy    608 HALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDN 667
             |:|::         :::   :|:|:|              |     ::|:|::|:|::||N
Db   1052 sdld--mssevdvggymsdg---dilgks------lrtd----dinsgymtdg 1089

Qy    668 QLGLHHTN----PEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDG 723
             :|:|::|     |:|:::|:|:||                         :|:|:|||:|G
Db   1090 glnlytrslnripdtatsrdi----iqrg-vhdvtvdadswdsssvvsglsdtldn 1141

Qy    724 AAYLLKDGEVKRLNINQSTSSI---------------------KHGTENVFSLPHVRNKPE- 763
             ::||| :::  :|:|:|::::|                     :||||:|:|:|:|:|:|
Db   1142 is------tddlnttssvssynitvpsrkntqlrtdsekrstdetwdspeelkkpex 1194

Qy    764 ---PGDA------LQGLNKD-DKAQAMAVIGVNK----YLALTEKGDIRSFQIKPGTQ 807
              |||        ::|::|| :||::|:|::|::|    ||||||||:|||:|:|||||Q
Db   1195 dfdshdgaggkwktvssglpedpekagqkagslsvsqtgswrrgmsagggapsrq-kagts 1253

Qy    808 QLERPAQT----LSREGISGELKDIHVDHKQNLYALTHEGEVFHQPREAWQNGAESSSWH 863
             :||         |||||||::|||:|:|:|:|:|||||:|:|:|||:||:|:|||||:|H
Db   1254 alktpgktddakasekg-kaplkgsslqrspsdagksdegkkppsgigrstatssfgf 1312

Qy    864 KLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTGSQTV 923
             |                ::|     :||   :     ::::        |:|:|:|:|V
Db   1313 k----------kpsgv--gssamitssg--------atitgsati 1338

Qy    924 FNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRK 968
             :|::                ||:::  :|::|:|:|K
Db   1339 -----gk-----ipksaaigksna-grktsldgsqngddvvlhvsskttlqy 1380

Qy    969 RAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDL 1028
             |          |||::|        :|:|::   :                       :L
Db   1381 r------slprpsksst---sgipgrgg 1407

Qy   1029 QSKLETLDLGEHGAEL---------LNDMKRFRDELEQSATRSVTVLGQHQGVL 1073
             :|::     ::|:::|          ::: :   ::|   |||:|:|:::|:|L
Db   1408 dsnvssksagattskireptkigsgrsspvtvngtdkekevavsdsesvslsgspkssp 1467

Qy   1074 KSNGEINSEFKPSPGK-------ALVQSFNVNRSGQDLSKSLQQAVHATP--PSAESKL 1123
             ||         :|||K       :|:|:     |:|::|:|:::|:||    |||:||L
Db   1468 tsasacqaglrqpgskypdiasptfrrlfgakaggksaapntegvksssvmpspsttl 1527

Qy   1124 -----QSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKT--------ALTKSRLILD 1166
                  :|:|   ::||::  :|   ||:|: |:|:|          |::|:|:|:D
Db   1528 argqslespssgtgsmgsag--glsgsssplfnk-psdlttdvislshslasspasvh 1582

Qy   1167 TVTIGELHELADKALVSDHKPDADQIKOLRQQFDTLREKRYESNPVKHYTDMGFTHNKA 1226
             :|:   ::       |||       |:|:|:|:   |||::|::|::|:|:|:|::|:A
Db   1583 sftsgglvwaanmsssssagsk------dt--psyqsmtslh------tsses 1620

Qy   1227 LEANYDAVKAFINAFKKEHHG--VNLTIRTVLESQGSAELAKKLKNTLLSLDSGESMSFS 1284
             |:|       :::      ::   :: | ||:||||:|:|:|||::|:||:|||:||:|S
Db   1621 id------lpishhgslgltgt-hevqsllmrtgsvrstl------sesmqld 1662

Qy   1285 RSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGGVSGNI 1344
             |:|||||||:|:|||||||:|:|:|||||:|||||||||||||||||:|:|:|||:|||I
Db   1663 rn------tlpkk-----------glrytpssrqangeeqkew----------- 1688

Qy   1345 MVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNS---LKFK 1401
             :::   :::     :||::|:  :|:|::|::|:|:|:|||||:|:||::|     LKFK
Db   1689 ------lrshstgglqdtgnqsp----lvsp----samsssaagkyhfsnlvsptn 1730

Qy   1402 LTEDELPG-----------FIHGLTHGTLTPAELLQKGIEH----------QMKQ 1435
             |         :::::     ::|:|:::|:::|:|::|:::          QMKQ
Db   1731 lsqfnlpgpsmmrsnsipaqdssfdlyddsqlcgsatsleerpraishsgsfrdsmeevh 1790

Qy   1436 GSKLTFSVDTSANLD----------LRAGINLNEDGSKPNGVTARVSAGLSASANL- 1481
             |:|                       |||:|:|::||||||||||:|:|||||||
Db   1791 gsslsvsstslystaeekahseqihklrrelvasqe---kvatitsqlsanahlv 1844

Qy   1482 AAGSRERSTTSGQFGSTTSASNNRPT-------FLNGVGAGANLTAALGVAHSSTHE 1531
             ::|               :|::|:||       ::|:|:||:|:|:||||:||:|:E
Db   1845 aafekslgnmtgrlgslmtaeqkeselielretiemlkaqnsaaq-aaigqalngpdhp 1903

Qy   1532 GKPVGIFPAFTSTNVSAALALDNRTSQSI-------SLELKRAEPVTSNDISELTSTLG 1583
             |          :::|:|||||:|:|:|          ||:||||:|||:|||:||:|:G
Db   1904 pkdlrirrqhssesvs---sinsatshssigsgndadskkkkknwvnsrg-selrssfk 1959

Qy   1584 KHFKDSATTKMLAA---LKELDDA---------KPAEQLHILQQHFSAK 1620
             |:::                :|::         |||:|||:|:|:|:|K
Db   1960 qafgkkkstkppsshsdieeltdsslpaspklphnagdcgsasmkpsqsasaicecteae 2019

Qy   1621 -DVVGDERYE-AVRNLKLVIRQQAADSHSMELGSAHSTTYNN-LSRINNDGIVELLHK 1677
              ::::       |:|::|:::|::|::|:|:|||:||::|: :|:|::|:|:|:|:|K
Db   1621 KHGTENVFSLPHVRNKPE------------------------------------- 763
```

```
Db   2020 aeiilqlkselrekelkltdirlea------lssahhldqireamnrmqne--ieilka 2070

Qy   1678 HFD-----AALP-----------ASSAKRLGEMMNN----DPALKDII------- 1705
               ||:|    ||      :||   :||       :|   :|
Db   2071 endrlkaetgntakptrppsessstsssssrqslgislnniteavssdilddagda 2130

Qy   1706 ------------KQLQSTPFSSASVSMELKDGLREQTEKAILDGKVGR-- 1741
                       :|: :| |||| :| :|||| ||
Db   2131 tghkdqrsvkiivsiskgygrakdqksqayiligsi----gvsgktkwdvldgviirrlf 2184

Qy   1742 EEVGVLFQDRNNLRVKSVSVSQ---SVSKSEGFNTPALLLGTSNSAAMSMERNIGTINF 1797
               |: |     |   ||       :|  |   :  ||  :|
Db   2185 keyvfridtsslglissdciasycigdliirshlevpell---pcgylvgdnniitvnl 2240

Qy   1798 KYGQDQNTPRRFTLE 1812
               | |:  |||
Db   2241 k-gveensldsfvfd 2254


RESULT 10
B41087

ID   B41087 standard; Protein; 2541 AA.
XX
AC   B41087;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human ORFX ORF851 polypeptide sequence SEQ ID NO:1702.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999; 99US-0127607.
PR   02-APR-1999; 99US-0127636.
PR   05-APR-1999; 99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   N-PSDB; C75296.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 1357; 5507pp; English.
XX
CC   C74446 to C77606 encode the proteins given in B40237 to B43397, which
CC   represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
```

```
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   anticonvulsant; antiarthritic; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 2541 AA;

Query Match             2.4%;  Score 224.5;  DB 21;  Length 2541;
Best Local Similarity  19.4%;  Pred. No. 5.6e-05;
Matches 447; Conservative 299; Mismatches 843; Indels 713; Gaps 113;

Qy     11 KAAVHTAAHNPVG---HG-VALQ--QGSSSSSPQN------------AAASLAAEGKNRGKMPR 56
          |||| |:   |||    ||  |||  |:||| |:|||
Db    428 kstvlqqqynrvgkvehgsvalpaimrsgasgppenfqvsgmppaqqqitsggqmhrghmpp 487

Qy     57 IHQPSTAADG------------ISAAHQQKKSFSLRGCLG---TKKFSRSAPQGQPGTTHSK- 103
          | |   :||              :||| ||::| ::|      ||  | :: :|||
Db    488 ltsaqqaltgtinssmgavqaaqatlddfdtlpplgqdaaskawrknkmdeskheihsqv 547

Qy    104 ------GATLRDLLARDDGETQHE--------AAAPDAARLTR-SGGVK--RRNMDDM--AGRP 148
                |:|| ::| ||  ||          |  :|||:|      ||||      :| :|   ||
Db    548 daitagtasvvnltagdpaetdytavgcavttissnltemsrgvkllaalledeggsgrp 607

Qy    149 MVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQH----SPPHIPG 203
          |:  :  |:|||  :|| |:  :| |  || | :  |::: ||:|      : |  |||
Db    608 llqaakglagavsellrsagpasaeprqnliqaagnvqqa-sgellqqigesdtdphfqd 666

Qy    204 SHHIKEEPVGSTSKATTAHADRVEIAQEDDSEFQQLHQQRLARERENPPQPPKLGVAT 263
          | |::|  |||||:|    ::|::|: ::|: || :|  :|   |::::|:: || ||
Db    667 almql-akavasaaaalvlkak--svaqrtedsglq--tqviaaatgcalstsqlvact 720

Qy    264 PISARFQPKLTA----------------VAESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAVT 310
          |:|||| |::|                 ||:|: | :|::|:|   || :|: : | ||:|||
Db    721 kvva--ptisspvcqeqlveagrlvakavkgcvsasqaatedgqllrgvgaaat--avt 775

Qy    311 LDKGKLQLAPDNPALNTLLKQTLGKDTOHYLAHHASSDGSQHLLLDNKGHLFDIKSTAT 370
          |:||: ||||:|  | |  |::    ||| | : :: |||                    :|:
Db    776 -----------------qalnell-------qhvkahatgagpagr------yd-qatdt 804

Qy    371 SYSVLHN-----SHPGEIKGK---LAQAGT--GSVSVDGKSGKISLGSGTQSHNKTMLS 419
          ||:|| |      ||| :|||    | ||    | |||: ||:||:|||                |||
Db    805 ilvtenifssmgdagemvrqarilaqatsdlvnaikadae-----gesdlensrklls 858

Qy    420 QP--GEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLS 476
          |:    |  ::: :     ||   | :|||      :  ||||  :| :|
Db    859 aakiladatakmv--eaakgaa----ahpd------seeqqqrlr 891

Qy    477 RQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQOVAILTDTPGRHKMSIMPS 536
          |  |: | | :  ||   :| |  : : |:|: : |:::   |::: ||| || | :
Db    892 eaaeg--lrmatnaaaq---naikklvqrle-haakqaaasatgtiaaaqhaas-apk 943

Qy    537 LDASPESHISLSLH-FADAHQGLLHGKSELEAQSVAISHGRLVVADSE-----GKLFSA 589
          | |||: | :|||  ||:  |:|||:|||  | : |: | ||||:|:     ||:|||
Db    944 asagpqpllvqsckavaeqipllvggvrgsqagpdspsaqlaliaasqsflqpggkmvaa 1003

Qy    590 A---IPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDHDHGQLNALVKN-- 637
          |     |  :|| : :| ::| |:||  |  |  ::   :|: ||| |:|||
Db   1004 akasvptiqd---qasam------------qlsqcaknlgtgtalaelr 1035
```

us-09-596-784-2.rag

Wed Jun 6 10:19:08 2001

```
Qy    638  --NFRQQHAC-PLGNDHQFHPGWNL------TDALVIDNQLGLHHTNPEPHEIL------D  683
                :|:   ||                   :||       ||:   | |
Db   1036  taaqkaqeacgplemdsalsvvqnlekdlqevkaaardgkl----kpipgetmekctqd 1090

Qy    684  MGH---------LGSLALQBGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDG  731
            :|:        |   | || :|   |       |  :
Db   1091  ignstkavssaiaqligeva--qgnenyagiaardvagg---lrslaqargvaaltsdp 1145

Qy    732  EVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDKAQAMAVIGVNKYL-- 789
            |  ||:| :|| |    ||  ||| :: || :|
Db   1146  avqaivldtasdvldkaslieeakkaaghpgdpesqqrlaqvakavtqa---lnrcvsc 1202

Qy    790  -ALTEKGD--IRSFQIKPGTQQLERPAQTLSREGISG----ELKDIHV 830
            ||  ||     || | | |  | || :||       |   |
Db   1203  lpggqrdvdnalravgdaskrllsdslppstgtge-aqsrlneaaaginqaatelvqasr 1261

Qy    831  DHKQNL-YALTHEGEVFHQPREAWQNGAESSSWHKLALPQSE----SKLKSLDMSHEH  883
            :|:|   || |  |  |      |           |        |  |  |:|
Db   1262  gtpqdlarasgrfggdfstflea---gvemag---gapsqedraqvvsnlkgismsssk 1314

Qy    884  -----KPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTV--------  923
                  |      |  | |:  |         |    |
Db   1315  lllaakalst--dpaapnlks---qlaaaar--avtdsinglltmctqapgqkecdn 1365

Qy    924  FNRLMQGVKGKVIPGSGLT-VKLSAQTG-----GMT  953
            |        |: |    :|  |  || |      |
Db   1366  alreletvrellenpvqpindmsyfgcldsvmenskvlgeamtgisqnakngnlpefgda 1425

Qy    954  GAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQ 1013
             :|   |                  |   |  |
Db   1426  istaskalcgfteaaaqaaylvgvsdp---------nsgagqglveptqfara------ 1470

Qy   1014  LDAHVRHNAPQPDLQSKLETLDLGEHG-----AELLNDMKRFRDELEQSATRSVTVLGQHQ 1070
            |          |         |         || |
Db   1471  ------ngaiqmacqs---igepgctqagvls----------aativakht 1502

Qy   1071  GVLKSNGEINS--EFKPSPGKALVQSF-NVNRSGQDLSKSLQ---------QAVHAT 1115
            |  |          |  |                  :|
Db   1503  salcnscrlasarttnptakrgfvqsakevanstanlvktikaldgafteenraqcraat 1562

Qy   1116  PPSAE-----------SKLQSMLGH----------FVSAGVDMSHQKGEIPLGR----- 1148
            |                |              ||
Db   1563  aplleavdnlsafasnpeifssipaqispegraamepivisaktmlesaggliqtaralav 1622

Qy   1149  -QRDPNDKTALT-KSRLILDTV------------TIGELH-ELA----------DKAKLVS 1184
            ||      |  |:||::|           ||  |:|           || |
Db   1623  nprdppswsvlaghsrtvsdsikkllitsmrdkapgglecetaiaalnsclrdldqaslaa 1682

Qy   1185  -------------------DH-----KPDADQ---IKQLRQQFD--T 1202
                               |       |  |      |:|  |
Db   1683  vsqqlapregisqealhtqmltavqeishlleplanaaraeasqlghkvsqnaqyfeplt 1742

Qy   1203  LREKRYESNPVKHYTDMGFTH--------NKALEANYDAVKAFINAFKKEHHGVNLTTRTVL 1256
            |  |:   | |                |  |       |  :|
Db   1743  laavgaasktlshpqqmalldqtktlaesalqllytakeagqnpkqah------tqeal 1796

Qy   1257  ES--QGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGIT 1314
            :|   |  |   :    |  |  ::|  :|     ||:|
Db   1797  eeavqmnteavedlttli----neaasaagvvggmvdsi-tqainqldegpngepegsf 1850

Qy   1315  LDRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTG----------------- 1358
             |  | :|   |         | |
Db   1851  vd--yqttmvtakaiavtvqe----------mvtksntspeelgplanqltsdygrlase 1899

Qy   1359  -KKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLT------1415
            | |    |:|  |   | || | |  |  :| | |  |
Db   1900  akpaavaaeneeigshikhrvq-elghgcaalvtkagalqcspsdaytkkel---------1950

Qy   1416  HGTLTPAELLQKGIEH---QMKQGSKLTFSVDTSAN------LDLRAGI-------NLNED 1460
            |   |   :|  | |    |    |   | :  :|       ||
Db   1951  ---iecarrvsekvshvlaalqagnrgtqacitaasavsgiialddttimfatagtlnre 2007

Qy   1461  GSKPNGVTARVSAGLSASANL---------------AAGSRERSTTSGQ----------F 1495
```

```
Db   2008  gte--tsadhregilktakvlvedtkvlvqnaagsgeklaqaaqssvatitrladvvkl 2064
            |||:|:     |                  :|        |
Qy   1496  GSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTH--EGKPVGIFPAFTSTNVSAALALD 1553
            ||| :::      |  |   |:  | |          ||  ||
Db   2065  gaaslgaedpetqvvlinavkdvakalgdlisatkaaagk-vgddpavwqlknsakvmvt 2123
            :|:  |     :     |   | |||       |
Qy   1554  NRTS---------------------------QSISLEL-----KRAEPVTSNDISELTSTLG 1583
            :|                                |     ||  || :| | | |
Db   2124  nvtsllktvkavedeatkgtraleattehirqelavfcspeppaktstpedfirmtkgi- 2182
             |  |         |  |      |   | |::      || | |:| | |
Qy   1584  KHFKDSATTKMLA------------ALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVR 1632
             | |:          |   |  | |   | : ||  |||
Db   2183  ----tmatakavaagnscrqedviatanlsrraiadmlrackeeaayhpevapdvrlralh 2238
            | ||                |:|
Qy   1633  NLKKLIV-----------------IRQQAADSHSMELGSASH- 1656
            ||:||                 ::|
Db   2239  ygrecangylelidhvlltlqkpspelkqgltghskrvagsvtellqaaeamkgtewvdp 2298
            ||  |   |  || | |  |  | | |  |   ||       |  |
Qy   1657  -STTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMNNDPALKDIIKQLQSTPFS  1714
            :|: |::   | ::|  |  | |:  |   |   :|   | |   | |
Db   2299  edptviaenellgaaaaieaaakkleqlkprakpkeadeslnfeeqileaaksi------ 2352
            |||  : |  |:   ||  |      |   | ||      |
Qy   1715  SASVSMELKDLREQTEKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQSVSKS------ 1768
             :|: |     |||           ||          |     |   |
Db   2353  aaatsalvkaasaqre-ivaqgkvgaipanal--ddgqwsqglisaarmvaaatnlce 2409
            |  | |        |  ||  |    |   |      |      |         |
Qy   1769  -----EGFNTPALLLGTSNSAAMSMERNIGTINFKYGGQDQNTPRRFTLEG-GIAQAN-- 1819
                   |  | | || |       |    |     ||      |      |
Db   2410  aanaavqghasqeklissakqvaastaqllvackvkadqdseamkrlqaagnavkrasdn 2469
                   | ||  :|  |  ||  |      |:  | |  | |
Qy   1820  --PQVASALTDLKKEGLEMK 1837
              ||   |    |  |  |
Db   2470  lvkaaqkaaafeeqenetvvk 2491
```

RESULT 11

Y85569
ID   Y85569 standard; Protein; 2385 AA.
XX
AC   Y85569;
XX
DT   07-JUL-2000 (first entry)
XX
DE   Human homologue of UNC-53 (Hs-UNC-53/3) sequence.
XX
KW   UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
KW   anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
KW   antisclerotic; antimetastatic; anti-arthritic; autoimmune disease.
XX
OS   Homo sapiens.
XX
FH
FT   Key   Location/Qualifiers
FT   Region   1..81
FT   /note= "this region can be replaced with the sequence
FT   shown in Y85570 to create a variant"
FT   Region   1326..1413
FT   /note= "this region is found to be absent when encoded by
FT   a variant cDNA"
FT   Region   1414..1427
FT   /note= "this region is found to be absent when encoded by
FT   a variant cDNA"
FT   Region   1703..1709
FT   /note= "this region is found to be absent when encoded by
FT   a variant cDNA"
FT   Region   1768..1788
FT   /note= "this region is found to be absent when encoded by
FT   a variant cDNA"
FT
XX
PN   WO9963080-A1.
XX
PD   09-DEC-1999.
XX

```
PF
XX
PR   02-JUN-1999;  99WO-EP03848.
PR   03-JUN-1998;  98GB-0011962.
XX
PA   (JANC )  JANSSEN PHARM NV.
XX
PI   Luyten WHML,  De Raeymaeker MC,  Geysen JJGH,  Bogaert TAOE;
PI   Maerten LJS,  Verhasselt P,  Van De Craen M;
XX
DR   WPI:  2000-116370/10.
DR   N-PSDB:  A07840.
XX
PT   Novel proteins and nucleic acids e.g. for treating neurodegeneration -
XX
PS   Claim 3; Fig 1f; 146pp; English.
XX
CC   The invention provides vertebrate (human) protein homologue of a UNC-53
CC   protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
CC   their plus ends. The UNC-53 sequences are used to promote neural
CC   regeneration, revascularization and wound healing; also for treating
CC   neurodegenerative disease, acute traumatic injury, fibrotic disease and
CC   autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
CC   polynucleotides can be used for recombinant production of the proteins,
CC   as a source of probes for detecting allelic variants and polymorphisms,
CC   for sequencing genomic DNA and for detecting UNC-53 expression, and as
CC   source of therapeutic antisense sequences. Cells that express the
CC   protein are used to identify regulators of cell shape, growth, motility
CC   and migration. They can also be used to identify proteins that are
CC   involved in signal transduction pathways also involving UNC-53, and to
CC   identify compounds that alter attachment of UNC-53 to microtubules. A
CC   target gene coupled to a UNC 53 encoding sequence may be used to deliver
CC   the target gene to a cellular microtubule or its plus ends. The present
CC   sequence represents the amino acid sequence of the third human homologue
CC   of UNC-53, designated hs-UNC-53/3.
XX
SQ   Sequence 2385 AA;
```

```
Query Match             2.3%;  Score 222;  DB 21;  Length 2385;
Best Local Similarity   18.5%;  Pred. No. 7.3e-05;
Matches  405;  Conservative  289;  Mismatches  736;  Indels  760;  Gaps 106;

Qy     22 VGHGVALQQGSSSSSPQNAAASLAAE----------GKNRGKMPRIHQP--------STA 63
Db    203 vthasppseasgaktgqdmqsslaaryatqshshsgiatsgkkptrlpgpsrvpaagsssk 262

Qy     64 ADGISAAHQQKKSFSL-------RGCLGTKKFSRSAPGQGPGTTHSKGATLRDLLARDDG 116
Db    263 vggasnlnrrsgsfnsidknkppnyangnekdsskgpqsssgv-----------ng 307

Qy    117 ETQHEAAA-----------PDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRH 166
Db    308 nvqppstagqppasaipspsaskpwrs---ksmnvkhsatstml------------ 348

Qy    167 QLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADR 226
Db    349 ----------tvkqsstatsptpss----drlk--ppvseq---vktapsqks--------- 383

Qy    227 VEIAQEDDDSEFQQLHQQRL-ARERENPPQPPKLGV----------------- 261
Db    384 ----------mlekfklvnartalrppqppssgpsdggkdddafsesgemegfnsgl 430

Qy    262 ---ATPISARFQPKLT------AVAESVLEGTD-------TTQSPLKPQ-S 295
Db    431 nsggstnsspkvspklappkagsknlsnkkssllqpkekeeknrdknkvctekpvkeekdq 490

Qy    296 MLKGSGAGVTPLAVTLDKG----------KLQLAPDN-----PPALNTLLKQTL----GKD 337
Db    491 vtemapktskiaslipkgsktaakkeslipsssgipkpgskvptvkqtispgstaske 550

Qy    338 TQHYLAHHASSDGSQHLLLDNKGHGHLFDIKSTATSYSVLHNSHPGEIKGKLA-----QAGTG 393
Db    551 sekfrttkgspsqs---------------lskpitmekasasccpaplegreagqaspsgsc 597

Qy    394 SVSVDGKSGKISLGSGT----------QSHNKTMLSQP--GEAHRSLLTGIWQHPAGAARPQ 443
Db    598 tmtvaqssgq-stgngavqlpqgqqhshpntatvapfiyrah------senegtalps 648

Qy    444 GESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSD-NKSSE 502
Db    649 adsc-tsptkmdlsysk-----------takqcleeisgepetrrmtvkniadlrqnle 697

Qy    503 KLVDKIK------SYS-VDQRGGVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQG 557
Db    698 etmsslrgtqishstlettfdstvttevngr------tipnltsrp--tpmtwrlgqacpr 750

Qy    558 LLHGKSELEAQSVAISH----GRLVVADSEGKLFS----AAIPKQGDGNELKMKAMPQ 607
Db    751 lqag---dapslgagyprsgtsrfihtdpsrfmyttplrraavsrlgnmsqidmsekas 806

Qy    608 HALDEHFGHDHQISGFFHDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDN 667
Db    807 sdld--mssevdvggymsdg---dilgks----------lrtd------dinsgymtdg 844

Qy    668 QLGLHHTN---PEPHEILDMGHLGSLALQEGKLHYFDQLITKGWTGAESDCKQLKKGLDG 723
Db    845 glnlytrslnripdtatsrdi------iqrg-vhdvtvdadswdgssvvssglsdtidn 896

Qy    724 AAYLLKDGEVKRLNINQSTSSI------------KHGTENVFSLPHVRNKPEP 764
Db    897 is------tddlnttsvssysnitvpsrkntqirtdsekrsttdetwdspeelkpee 949

Qy    765 -----GDA----LQGLNKD-DKAQAMAVIGVNK----YLALTEKGDIRSFQIKPGTQ 807
Db    950 dfdshgdaggkwktvssglpedpekagqkaslsvsgtgswrrgmsaggapsrq-kagts 1008

Qy    808 QLERPAQT---LSREGISGELKDIHVDHKQNLYALTHEGEVFHQPREAWQNGAESSWH 863
Db   1009 alktpgktddakasekg-kaplkqsslqzspsdagkssgdegkkppsgirstatssfgf 1067

Qy    864 KLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTV 923
Db   1068 k-----------------------kpsgv--gssamitssg-----atitsgsati 1093

Qy    924 FNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRK----------VSSKFSERI 968
Db   1094 k-----------------ipksaaiggksna-grktsldgsqnqddvvlhvssktt1qy 1135

Qy    969 RAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDL 1028
Db   1136 r---------slprpsksst---sgipgrgg-----------hrsstssi 1162

Qy   1029 QSKLETLDLGEHGAEL-----------LNDMKRFRDELEQSATRSVTVLGQHOGVL 1073
Db   1163 dsnvsskagattsklreptkigsgrsspvtvnqtdkekekvavsdsesvslsgspkssp 1222

Qy   1074 KSNGEINSEFKPSPGK----------ALVQSFNVNRSGQDLSKSLQQAVHATP--PSAESKL 1123
Db   1223 tsasacgaqglrqpgskypdiasptfrrlfgakagqksasapntegvksssvmpspsttl 1282

Qy   1124 -------QSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKT-----ALTKSRLILD 1166
Db   1283 arqgslespssgtgsmgsag--glsgsssplfnk--psdlttdvislshslasspasvh 1337

Qy   1167 TVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKA 1226
Db   1338 sftsggivwaannsssssagsk----------dt--psyqsmtslh------tsses 1375

Qy   1227 LEANYDAVKAFINAFKKEHHG--VNLTRTVLESQGSAELAKKLKNTLLSLDSGESMSFS 1284
Db   1376 id----------lpishhgslsgltgt-hevgsllmrtgsvrstl-----sesmqld 1417

Qy   1285 RSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGVSGNI 1344
Db   1418 rn----------tlpkk----------------glrytpssrqanqeegkew------ 1443
```

```
PR   09-OCT-1998;  98US-0103796.
PR   25-FEB-1999;  99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; Z53858.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 851; 1453pp; English.
XX
CC   253015 to 254536, 254577 to 254615, and Y74253 to Y75941 represent
CC   novel Neisseria meningitis and N. gonorrheae polynucleotides and
CC   polypeptides. 254537 to 254576 and 254616 to 254473 represent PCR
CC   primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists, or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence  2048 AA;

Query Match            2.3%;   Score 221.5;  DB 21;  Length 2048;
Best Local Similarity  18.5%;  Pred. No. 6.2e-05;
Matches  378;  Conservative  255;  Mismatches  763;  Indels  645;  Gaps  96;

Qy   124 APDA-ARLTRSGGVKRRNMDDM------------AGRPMVKGGSGEDKVPTQOKRHQLNFG 172
                    :|      :        |||                  :: ||::: | :  | |::
Db   277 aidangrlvnsgtmaaanvqdmnntaehkvnirsqafensgtavsqggtqihsqsiqntg 336

Qy   173 QMRQTMLSKMAHPASAN----AGDRL----QHSPPHIPGSHHEIKE----EPVG 214
             :: |        |       |||:         : | ||:: |        ::
Db   337 kllsagtedlavsgslnngneiatngqllihdgqgstvvidntngtiqsgrdvaiqaks 396

Qy   215 STSKATTAHADRVEIAQEDDDSEFQOLHQQRLARERENPPQPPKLGVATPISARFQPKLT 274
              :     ::   ::|:  |                                |
Db   397 lsngtlaadnklidialqdd------------------f 417

Qy   275 AVAESVLEGTDTTQSPLKFQSMLKGSGAGVTPLAVTLDKGK-LQLAP--DNPPALNTLL 330
              : | |  |         |         |        : |       |
Db   418 yverkivagnels--lstrgslknsh------tlqagkrirkannldnavqgniqs 466

Qy   331 KQTLGKDTQHYLAHHASSDGSQHLL------------LDNKGHLFDIK 366
              ||       |:  |     |:|             :||:::: |:
Db   467 ggtdigtqhnltnrglidgqqtkiqagqmnnigtgriygdniaiaatrldnq---den 522

Qy   367 STATSYSVLHNSHPG----------------EIKGKLAQAGTGSVSVD 398
                      :   |                   ::
Db   523 gtgaaiaarenlnlgieqlnnrensliysgndmavggaldtndqatgkaqrihnagaiie 582

Qy   399 GKSGKISLG---SGTQSHNKTMLSQPGE----AHRSLLTGIWQHPAG--AARP 442
              ::|   |:|                       :
Db   583 a-agkmrlgveklhntnehlktqlvetgreridyeafgrhellregt-qhelgwfvynn 640

Qy   443 QGESIRLHDDKIHILHPELGVWQSADKD--THSQLSRQADGKLYALKDNRTLQNLSDNK 499
               ::  :   |  : | :          :|    ::  | :
Db   641 esdhlrtpdgvah-----enwhkydyekvtqetqvtgtapakila------ 680

Qy   500 SSERLVDKIKSYSVDQR----GQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAH 555
              | :: : |  :  :          || ::     ||   ||
```

```
Qy  1345 MVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNS---LKFK 1401
              :::|                    :|   |||           :|||   : ::
Db  1444 ------------lrshstgglqdtgnqsp-----lvsp--samsssaagkyhfsnlvsptn 1485

Qy  1402 LTEDELPG----------FIHGLTHGTLTPAELLQKGIEH--------QMKQ 1435
            |:: |||            |  : ||                        :|
Db  1486 lsqfnlpgpsmmrsnsipaqdsfdlyddsqlcgsatsleerpraishsgsfrdsmeevh 1545

Qy  1436 GSKLTFSVDTSANLD--------LRAGINLNEDGSKPNGVTARVSAGLSASANL 1481
             | || |  ::|           ||:      :  :|    :::
Db  1546 gsslslvsststslystaeekahseqihklrrelvasqe-----kvatltsqisanahlv 1599

Qy  1482 AAGSRERSTTSGQFGSTTSASNNRPT---------FLNGVGAGANLTAALGVAHSSTHE 1531
             :::  : :  :   :: ::                 ||| |        :: :
Db  1600 aafeksignmtgrlqsltmtaeqkeselielretiemlkaqnsaaq-aaiqgalngpdhp 1658

Qy  1532 GKPVGIFPAFTSTNVSAALALDNRTSQSI-------SLELKRAEPVTSNDISELTSTLG 1583
             | : | :    ::||                   ||:                :: | |
Db  1659 pkdlrlirrqhssesvs--sinsatshssigsgndadskkkkknwvnsrg-selrssfk 1714

Qy  1584 KHFKDSATTKMLAA--LKELDDA----------KPAEQLH 1611
             | :                |::              || |:
Db  1715 qafgkkkstkppsshsdieeltdsslpaspklphnagdcgsasmkpsqsasasplvwppk 1774

Qy  1612 ------ILQQHFSAKDVVGDERYEAVRNLKKLVIRQQ--AADSHSMELGSASHSTTYNN 1662
                   :||   | : :: :  ::  :|:   |:::
Db  1775 krqngpviykhrsricecteaeailiqlkselrekelkltdirlealssahhldqirea 1834

Qy  1663 LSRINNDGIVELLHKHFD---------AALP-------ASSAKRIGEMMNN----- 1697
                :|   | :       |            ||       |||| |:|:::
Db  1835 mnrmqne-leilkaendrlkaetgntakptrppsessstssssssrgslglslnnlnit 1892

Qy  1698 DPALKDII------------KQLQSTPFSSASVSMELKDGLRE 1728
             : |:                ||  ||       :|:| | |
Db  1893 eavssdilddagdatghkdgrsvkliivsiskgygrakdqksqayliqsi-----gvsg 1946

Qy  1729 QTEKAILDGKVGR--EEVGVLFQDRNNLRVKSVSVSQ---SVSKSEGFNTPALLGTSN 1782
              |::    |  |:     | :| : |   :  | :|      |::::   |  ::|
Db  1947 ktkwdvldgvirrlfkeyvfridtstsiglssdciasycigdlirshnlevpell----p 2002

Qy  1783 SAAMSMERNIGTINFKYGQDQNTPRRFTLE 1812
             : | :: | :|:| :: :|:::
Db  2003 cgylvgdnnlitvnlk-gveenslsfvfd 2031
```

```
RESULT  12
Y75096
ID   Y75096 standard; Protein; 2048 AA.
XX
AC   Y75096;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrhoeae ORF 564 protein sequence SEQ ID NO:1666.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;  99WO-US09346.
XX
PR   01-MAY-1998;  98US-0083758.
PR   31-JUL-1998;  98US-0094869.
PR   02-SEP-1998;  98US-0098994.
PR   02-SEP-1998;  98US-0099062.
PR   09-OCT-1998;  98US-0103749.
PR   09-OCT-1998;  98US-0103794.
```

```
Db   681 gsdliidskavfnsdsriiaggqllvgtekdglhneq-------------------       717

Qy   556 QGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFG   615
             :.    .    .   ::.    . .
Db   718 ------------tfgekkvfsengklhnywrarrkghdetghreq-nytlpeeit       759

Qy   616 HDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHT-   674
              : .:.   :   ::    .::
Db   760 rdislgsfayeshsk-alsrhapsgtelpqsnrd-----nirtak-sngislpytp      809

Qy   675 ---NPEPHEIL-------DMGH----------------LGSLALQEGKLH------Y      699
               :       :||       :::  :
Db   810 nsftplpgsslyiinpankgylvetdprfanyrqwlgsdymlgslkldpnnlhkrlqdgy   869

Qy   700 FDQ--------LT-KGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNIN---------     739
             :      :::       .     ::.:.  ::
Db   870 yeqrlineqiaeltghrrldgyqndeeqfkalmdnga-----taarsmnlsvgialsaeq   924

Qy   740 --QSTSSI------------KHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVN   786
               :.::           ::     .::.   .
Db   925 aaqltsdivwlvqkevklpdggtqtvl-mpqvyvrvkng----gi--dgkgalls--gsn   975

Qy   787 KYLALTEKGDIRSPQIKPGTQQLERPAQTLSREGISGEL----------KDIH------VD   831
              :.  :   ::  .    ::  .       .:|:||        .::|
Db   976 tqinvs--gslknsgtiagrnaliintdtl--dniggrihaqksavtatqdinniggils  1031

Qy   832 HKQNLYALTHEGEVFHQPREA--WQNGAESSSW-HKLALPQSESKLIKSLDMSHEHKPIAT   888
              :..   .:::  :  :::    ::|:||     :::
Db  1032 aeqtl--llnagnninngstakssgnaggsstyldrmagiyitgkekgvlaagaqkdini  1089

Qy   889 FEDGSQHQLKAGGWHAYAAAPERGPLAVGTSGSQTVF----NRLMQGVKGKVIPGSGLTVK   944
              :.    ::||     ::.  :::       .:::     :.
Db  1090 iaggisnqsdqggqtrlqagrdinldtvqtqkyqeihfdadnhtirgstnev-gssiqtk  1147

Qy   945 LSAQTGGWTGAEGRKVSSKFSE------RIRAYAFNP-TMSTPRPIKNAAYATQHGWQGR   997
              :.:  |||      :::         :.:  . :    :::  .
Db  1148 ----gdvtllsgnnlnakaaevgsakgtlavyaknditissgihagqvddaskh--tgr  1200

Qy   998 EGLKPLYEMQGALIKQLDAHNVRHNAPQPD-LQSKLETLDLGEHGAELLNDMKRFRDELE  1056
             ::|      .    :.       :|       :.:|
Db  1201 sg------ggnklvitdkaqshhetaqsstfegkqvvlqag-------nd--------  1237

Qy  1057 QSATRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQODLSKSLQQAVHATP  1116
              :.:    .:    ::       ::|    .       .::|
Db  1238 -------anilgsn--visdng------------trigagnhvrigttqtqsqsetyhqtq  1277

Qy  1117 PSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHEL  1176
              ::      :::       :::|    :::       .:::
Db  1278 ksg----------lmsagifftigsktntqenqsqsnehtgstvgslkgdttivaskhye  1327

Qy  1177 ADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFT------------  1222
              ::.     :::   :.       ::.:|
Db  1328 qtgsnvss--pegnnllstqsmdigaaqnqlnskttqtyeqkgltvgiqfaryrfgtts  1384

Qy  1223 ------HNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAK--KLKNTL  1272
                   :::     .::      ::|         ::.    .::
Db  1385 dcrstqsskqvggskndrvnamaaa-------nagwqayqtgkgaqnlangttnakqvs  1436

Qy  1273 LSLDSGESMS------------FSRSYIGGGVSTVFVPTLSKKVPVPVI------PGAGI  1313
              :|    :::            :.:   :::        .::        ::|
Db  1437 isitygeqnrqttqvqanqagasqiqaggkttlycrrcgeqsninitgsvsgragtgl  1496

Qy  1314 TLDRAYNLSFSRTS------------GGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKK  1360
              ::.       :::            ::|       :::
Db  1497 iadkqihlqsaeqsntersqnksagwnagaavsfgqgygwslg---vaaggnv-----gkg  1548

Qy  1361 TSAGNASDWLSAKHKISPD------------LRIGAAVSG-----------------  1388
              ::.       :::            ::|
Db  1549 ygyg--dsvthrhshigdkgsqtliqsggdtiikgaqvrgkgvqvnaknlsigsvqdre  1605

Qy  1389 -LKFKLTEDELPGFIHGLTHGTLTPAELLQKGIE-HQMKQGSKLT  1440
              :::       :::|
Db  1606 tyqskqqnagaqvtvgygfsasgdysqgskira-dhasvteqsgiyagedgyqikvgn----  1661
```

```
Qy  1441 FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS  1500
              :.:    :::   .     ::|
Db  1662 ----------htglkggiitssqsakdkgkn-rfstgtlagsdiqnysqyeqksfg-lgasva  1712

Qy  1501 AS--------NNRP------TFLNG----VGAGANLTAALGVAHSSTHEGKPVGIFPA  1540
             ::        :::       ::||
Db  1713 vsgktlgqgaknkpqdkhltsiadkngassvygsdsdsqssitksgintpkniqi--  1769

Qy  1541 FTSTNVSAALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKE  1600
              :::    :::       .     ::|
Db  1770 ---tdeaaqirltgkiaaqtkadi--------dtnvttdtaerh-------sgslkn  1808

Qy  1601 LDDAKPAEQLHILQ---QHFSAKDVVGDERYEAVRNLKKLVIRQQAADSHSME-LGSAS  1655
              :.:       .:::   :::
Db  1809 ifdkdrvqseldlqrtvsqdfskn--vqqtnteinqhidklkadkeaaetaaaelangd  1866

Qy  1656 HSTTVNNLSRINNDGIVELLKHFDAALPASSAKRLGEMMNN------DPALKDI-IKQ  1707
              :::         :::
Db  1867 metakrk--------------aheaqdaaakadnvqq-gkvilnmlasgqlaeptqsgagiaa  1913

Qy  1708 LQSTPFSSASVSMELKDGLREQTEKAILDGKVGREEVGVLFQDRNNLRVKSVSV-SQSVS  1766
              :::         :::|
Db  1914 ataspdvsyaiqqhfkd---------laggnangkltasqetahvlahavlgaava  1960

Qy  1767 KSEGFNTPALLLGTSNSAAMSMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASAL  1826
              :::|        .:::      ::|         :::
Db  1961 aaxgnnapagalgaggseaaa--piig--kwlyqkgdg------gslnaeeketvsai  2008

Qy  1827 T  1827
             :
Db  2009 t  2009
```

```
RESULT  13
R05041
ID   R05041 standard; protein; 3647 AA.
XX
AC   R05041;
XX
DT   02-OCT-1990 (first entry)
XX
DE   Filamentous haemagglutinin A.
XX
KW   Filamentous haemagglutinin; fhaB; whooping cough; vaccine.
OS   Bordetella pertussis.
XX
FH   Key              Location/Qualifiers
FT   Region           1..1069
FT                    /label=N-terminal
FT                    /note="highly basic; 98 kD"
FT   Region           1074..3647
FT                    /label=C-terminal
FT                    /note="acidic and hydrophilic; 140 kD"
FT   Cleavage-site    1069..1073
FT                    /label=trypsin-like proteolytic cleavage site
FT                    /note="confimed by SDS-PAGE peptide analysis."
FT   Binding-site     1097..1099
FT                    /label=cell recognition site
FT                    /note="surface exposed, hydrophilic, antigenic"
FT   Binding-site     2651..2653
FT                    /label=cell recognition site
FT   Region           2000..2700
FT                    /label=high concn. of charged residues
FT   Region           1800..2000
FT                    /label=hydrophobic region
FT   Region           2400..2500
FT                    /label=hydrophobic region
FT   Domain           44..69
FT                    /label=transmembrane helix
FT   Region           52..69
FT                    /label=transmembrane segment of helix
```

```
FT   Region          406..498
FT                   /label=repeat region
FT                   /note="ABABA repeat pattern"
XX
PN   WO9004641-A.
XX
PD   03-MAY-1990.
XX
PF   20-OCT-1989; 89WO-US04732.
XX
PR   27-OCT-1988; 88US-0263648.
XX
PA   (STRD ) LELAND STANFORD JUNIOR UNIVERSITY.
XX
PI   Relman DA, Domenighini M, Rappuoli R, Falkow S;
XX
DR   WPI; 1990-164024/21.
DR   N-PSDB; Q04668.
XX
PT   Nucleic acid sequences encoding Bordetella pertussis fhaB gene -
PT   and peptide cross-reactive with filamentous haemagglutinin of b
PT   pertussis, useful for diagnosis and treatment of whooping cough.
XX
PS   Disclosure: ; lpp; English.
XX
CC   The sequence encodes a protein of calculated MW 368 kD with a Ala/Gly-
CC   rich compsn.(27%) which is nearly identical to published chemical
CC   analysis of AA compsn. It has an isoelectric point of 6.79.
CC   DNA encoding the protein and recombinant peptides produced from it are
CC   useful diagnostically and therapeutically against pertussis. Dosage is
CC   pref. 25-75 microg/kg (single dose).
XX
SQ   Sequence    3647 AA;

Query Match       2.2%;    Score 210;  DB 11;  Length 3647;
Best Local Similarity   19.6%;   Pred. No. 0.00081;
Matches  439;  Conservative 267;  Mismatches 807;  Indels 724;  Gaps 115;

Qy      5 SLGTEHKAAVHTAAHNPVGHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQPSTAA 64
Db   1379 sldikkggaqvtvagryaehgevsiqqdytvsad-aialaaqvtqrg----------ga 1426

Qy     65 DGISAAHQQKKSFSLRGCLGTKKFSRSAPGGQPGTH-SKGATLRDLLARDDGETQHEAA 123
Db   1427 anitsrhdtrifsnkir-lmqplqvnaggpvsntgnlkvregvtv--taasfdnetgaevm 1483

Qy    124 APDAARLTRSGGVKRRNMDDMAGRPMVK----------- GGSG 155
Db   1484 a-ksatltsgaarn------agkmqvkeaativaasvsnpqtftagkditvtsrggfdn 1536

Qy    156 EDKVPTQQ---KRHQLNN-----------FGQMRQTMLSKMAHPASANAGDRLQHS 197
Db   1537 egkmesnkdivikteqfsngrvldakhdltvtasgqadnrgslkaghdftvqa-qridns 1595

Qy    198 PPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQOLHQQRLARERENPPQPP 257
Db   1596 --------------dihilnsaklentgrvda 1639

Qy    258 KLGVATPISARFQPKLTAVAE----SVLEGTD---TTQSPLKP--QSMLKGSGAGVTP 306
Db   1640 rndialdv-adftntgslyaehdatltlaqgtqrdlvvdqdhilpvaegtlrvkakslt 1698

Qy    307 LAVTLDKGKL---QLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHL 362
Db   1699 eietgnpgsllaevgenidnkqai------vvgkdltlssa-hgnvaneanallwaagel 1751

Qy    363 FDIKSTATSYSVLHNSHPGEIKGKLAQAG----TGSVSVDGKSGKISLG----SGTQS 412
Db   1752 -----tvkaqnitnk-----raalieaggnarltaavallnklgrliragedmhldapri 1800

Qy    413 HNKTMLSQPGEAHRS------------------LLTGIWQHPAGAARP--------- 442
Db   1801 entakls--gevqrkgvqdvgggehgrwsgigvynywlragngkkagtiaapwyggdlta 1858

Qy    443 ------QGESIRLH-----DDKIHILHPELGVWQSADKDTHSQLSRQADGK------ 482
Db   1859 eqsllevgkdlylnagarkdehrhllne--gvlqagg--hghiggdvdnrsvvrtvsam 1913

Qy    483 -----LYALKDNRTLQNLSDN-KSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKM 531
Db   1914 eyfktplpvsltaldnraglspatwnfqstyelld---ylldqnryeylwglyptytew 1969

Qy    532 SI---------MPSLDASPESHISLSLHFADAHQG-----LLHG--KS 563
Db   1970 svntlknldlgyqakpaptappmpkape--ldlrghtlesaegrkifgeykklqgeyeka 2027

Qy    564 ELEAQSV-AISHGRLVVADSEGKLFSAAI---PKQGDGNELKMKAMPQHALDEHFGH 616
Db   2028 kmavqaveaygeatrrvhdqlgrygkaiggmdaetkevdg-----------iiqefaa 2075

Qy    617 DHQISGFFHDDHGQLNA---LVKNNFRQQ-HACPLGNDHQFHPGWNLTDALVIDNQLG- 670
Db   2076 dlrtvyakqadqatidaetdkvaqryksqidavrl---qaiqpg-rvtlakalsaalgad 2131

Qy    671 ---LHHTN-----------------PEPHEILDMGHLGSLALQEGKLHYFDQLTKGW 707
Db   2132 wralghsqlmqrwkdfkagkrgaeiafypkeqtvlaag--agltlsngaih------nge 2183

Qy    708 TGAESDCKQ--LKKG--------LDGAAYLLKD-GEVKRLNINQSTSSIKHGTENVFSLPHV 758
Db   2184 naaqnrgpeglkigahsatsvsgsfdalrdvglekrldiddalaa----vlvnphi 2236

Qy    759 --RNKPE--------------PGDALQGLNKDKAQAMAVI 783
Db   2237 ftrigaqtsladgaapalargarqapetdgmvdarglgsadalaslasldaaqglevs 2296

Qy    784 GVNKYLALTEKGDIRSFQIKPGTQOLERPAQTLSREGISGELKDIHVDHKQNLYALTHEG 843
Db   2297 g-rrnaqvadag------lagpsavaapavgaadvgvepvtgd-qvdhpvvavgl----- 2343

Qy    844 EVFHQPREAWQNGAESSSWHKLALPQS-ESKLKSLDMSHEHKPIATFEDGSQHLKAGG 901
Db   2344 --------eqpvatvrvappa-----valrplfetrikfidqskfy-------gsryffeqig 2387

Qy    902 WHAYAAPERGPLAVGTSGSQOTVF----NRLMQGVKGKV-IPGSGLTVKLSAQTGGMTGA 955
Db   2388 y---kpdraarvagdnyfdttlvreqvrralggyyesrlpvrgvalvakimdsagtvgka 2443

Qy    956 EGRKV------SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQG 1008
Db   2444 lglkvgvaptaggqlkqadrdfvwydtvidqgkvlaprlyte--atrggitdqyaggg 2500

Qy   1009 AL----IKQLDAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSV 1063
Db   2501 vaptaqqlkqadrdfvwy--------vdtvidqgkvlap---rlyleatrggi 2543

Qy   1064 TVLGQHQGVLKSNGEINSEFKPSPGKALVQS---FNVNRSGQDLSKSLQQAVHATPPSAE 1120
Db   2544 t--dqyag---------ggaliasgqdvtvntdghdvs-----svngligqrs 2580

Qy   1121 SKLQSMLGHFVSAG------VDMSHQKGEIPLGRQRPDNDKTALTKSRLILDT 1167
Db   2581 vkvdagkgkvvvadskgagggieaddevdvsgrdiglieggklrgkd----vrlkadt 2633

Qy   1168 VTIGELHELADKAKLVS--DHKPDAD------QIKQLRQQFDTLREKR--- 1207
Db   2634 vkvatsmryddkgrlaargdaldagggqlhieakrletagatlkggkvkldvddvklgg 2693

Qy   1208 -YESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTV----LESQG 1260
Db   2694 vyeag--ssyenksstplgslfailssttetnqsahanhygtrieagtlegkmqnleieg 2751

Qy   1261 SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVIPGAGITLDRAYN 1320
Db   2752 gsvda--aitdlsvrdarfkaaadfahaehekdvrqlslgakv-----gag-----gye 2799
```

```
PA   (DOME/) DOMENIGHINI M.
PA   (RAPP/) RAPPUOLI R.
PA   (FALK/) FALKOW S.
XX
PI   Falkow S, Domenighini M, Relman DA, Rappuoli R;
XX
DR   WPI; 2000-269872/23.
DR   N-PSDB; A10263.
XX
PT   New isolated nucleic acid encoding the filamentous hemagglutinin of
PT   Bordetella pertussis, useful for diagnosis, treatment and prevention of
PT   whooping cough -
XX
PS   Disclosure; Columns 11-28; 17pp; English.
XX
CC   This sequence represents filamentous haemagglutinin (FHA) from
CC   Bordetella pertussis. Bordetella pertussis is a small Gram negative
CC   bacillus found only in humans and which causes whooping cough (pertussis)
CC   in children. Filamentous haemagglutinin is a 368 kD surface-
CC   associated protein which is thought to be one of the most important
CC   factors in mediating the adhesion of the bacterium to a eukaryotic
CC   (human) cell. FHA also stimulates an immune response in humans following
CC   disease, and has been shown to act as an immunoprotective antigen in an
CC   animal model. The FHA gene, or its expression products, are used for
CC   prevention or treatment of pertussis, particularly in vaccines. It may
CC   also be used to increase the amount of FHA in live or dead Bordetella
CC   pertussis or other organisms. Fragments of the gene may be used as
CC   diagnostic hybridisation probes or as antisense modulators to reduce the
CC   infectivity of Bordetella pertussis. FHA, or fragments thereof, may be
CC   used in immunoassays to detect anti-FHA antibodies, and to raise
CC   antibodies for the diagnosis, prevention of treatment of whooping cough.
XX
SQ   Sequence   3596 AA.

Query Match          2.2%;  Score 207.5;  DB 21;  Length 3596;
Best Local Similarity 19.3%;  Pred. No. 0.0011;
Matches 427;  Conservative 268;  Mismatches 792;  Indels 725;  Gaps 114;

Qy     5 SLGTEHKAAVHTAAHNPVGHGVALQQGSSSSSPQNAAASLAEGKNRGKMPRIHQPSTAA 64
           |   :|| :| || | |  || | ::||| :
Db  1379 sldikkggaqrtvagryaehgevsiggdytvsad--aialaaqvtqrg--------ga 1426

Qy    65 DGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTH-SKGATLRDLLARDDGETQHEAA 123
           :| ||| | ::| | ::   |   ||| | : || |:|
Db  1427 anltsrhdtrfsnkir-lmgplqvnaggpvsntgnlkvregvtv--taasfdnetgaevm 1483

Qy   124 APDAARLTRSGGVKRRNMDDMAGRPMVK----------------------GGSG 155
          | :|  |:|| |:   |  | :
Db  1484 a-ksatlttsgaarn------agkmqvkeaativaasvsnpgtftagkditvtsrggfdn 1536

Qy   156 EDKVPTQQ----KRHQLNN-----------------------FGOMRQTMLSKMAHPASANAGDRLQHS 197
          | |:| :   | |:||
Db  1537 egkmesnkdivikteqfsngrvldakhdltvtasggadnrgslkaghdftvqa-qridns 1595

Qy   198 PPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQOLHQQRLARERENPPQPP 257
Db  1596 gtmaaghdatlkaphlrntgvvagh----------------dihiinsaklentgrvda 1639

Qy   258 KLGVATPISARFQPKLTAVAE---SVLEGTD----TTQSPLKP-QSMLKGSGAGVTP 306
Db  1640 rndialdv-adftntgslyaehdatltlaagtgrdlvvdqdhilpvaegtirvkaksltt 1698

Qy   307 LAVTLDKGKL---QLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHL 362
Db  1699 eietgnpgsliaevgenidnkqai---------vvgkdltlssa-hgnvaneanallwaagel 1751

Qy   363 FDIKSTATSYSVLHNSHPGEIKGKLAQAG----TGSVSVDGKSGKISLG----SGTQS 412
Db  1752 ----tvkaqnitnk------raalieaggnarltaavallnklgriragedmhldapri 1800

Qy   413 HNKTMLSQPGEAHRS 442
          | |  ||  | ||
```

```
Qy  1321 LSFSRTS-GGLNVSFGRDGGVSGNIMVATGHDVMPYMT----------------GKKTSAG 1364
           |  || | |   | | ||  | ::
Db  2800 agfslgsesgleahagr--gmtagaevkvgyrasheqsseteksyrnanlnfgggsveag 2857

Qy  1365 NASDWLSAKHKISPDL---RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLT------ 1415
           |: : |  :: | | :  ::   |
Db  2858 nvldigga------dinrnryggaakgn-agt-eealrmrakkvestkyvseqtsqssgw 2909

Qy  1416 ----------HGTLTPAELLQKGIEHQMKQGSKL---------TFSVDTSANL 1449
                    |  ||  : :|  ::
Db  2910 svevastasarsslltaatrlgdsvaqnvedgreirgelmaaqvaaeatqlvtadtaa-v 2968

Qy  1450 DLRAGINLNEDGSKPNGVTA-------------------RVSAGLSASAN 1480
           |  |    :    :|         ::  |  |  |
Db  2969 alsagisadfdsshsrstsqntqylggnlsieategdatlvgakfgggdqvslkaaksvn 3028

Qy  1481 LAAGSR--ERSTTSGQFGSTTSASNNRPTFLNGVAGANLTAALGVAHSSTHE-GKPVGI 1537
           |  ||    || : ::|  : | : |: : | :|   |     ::| |:|
Db  3029 lmaaestfesyseshfhasadanlganavggavglg-ltagmgtshqitnetgk----- 3082

Qy  1538 FPAFTSTNVSAA-LALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLA 1596
             | :|:::  :|  :|:     |||
Db  3083 --tyagtsvdaanvsid--agkdlnlsgsrvr---------gkhvlvdvegdina 3124

Qy  1597 ALKELDDAKPAEQLHILQQHFSAKDVVGDER-YE-------AVRNLKKLVIRQQA 1643
            : ||                  :|   ||
Db  3125 tskq---------------dernynssgggwdasagvaiqn-rtlv-------- 3154

Qy  1644 ADSHSMELGSA--SHSTTYNNLSRINNDGIVELL-------HKHFDAALPASSAKRLGEM 1694
            |: :   | :| :|:: || || || ||     |||
Db  3155 ----apvgsagfnfntehdn-srltndgaagvvasdgltghvkgdanltgatiadlsgk 3208

Qy  1695 MNNDPALKDIIKQLQLQSTPFSSASVSMELKDGLREQTEKAILDGKVGREEVGVLFQDRNNL 1754
              |    |     :   |          :|  :  :|
Db  3209 gn-----lk--------vdgavnaqnlkd-yrdk-----dggsglnvg--------- 3238

Qy  1755 RVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYGQDQNTPRRFTLEGG 1814
           :   : |  | :|   |       |:|    :        ||
Db  3239 -issttlaptvgvafg------rvagedyqaeqratidvg--------qtkdparlqvgggg 3284

Qy  1815 IAQANPQVASALTDLKK 1831
           : | |   |:|:
Db  3285 vkgtlnqdaaqatvvqr 3301
```

```
RESULT  14
Y87407
ID   Y87407 standard; Protein; 3596 AA.
XX
AC   Y87407;
XX
DT   03-JUL-2000 (first entry)
XX
DE   Bordetella pertussis filamentous haemagglutinin (FHA).
XX
KW   Filamentous haemagglutinin; FHA; fhaB gene; surface-associated;
KW   cellular adhesion; antigenic; immunogenic; whooping cough; pertussis;
KW   diagnosis; prophylaxis; vaccine.
XX
OS   Bordetella pertussis.
XX
PN   US6036960-A.
XX
PD   14-MAR-2000.
XX
PF   01-SEP-1994; 94US-0299941.
XX
PR   13-NOV-1989; 89US-0436297.
PR   10-AUG-1992; 92US-0928619.
PR   27-OCT-1988; 88US-0263648.
XX
PA   (RELM/) RELMAN D A.
```

```
Db   1801  entakls--gevqrkgvqvdvggehgrwsgigvynvwlragngkkagtiaapwyggdlta  1858
Qy    443  ------QGESIRLH---DDKIHILHPELGVWQSADKDTHSQLSRQADGK----------   482
           ::|   |:                          |:  |::|  : |:
Db   1859  egslievgkdlylnagarkdehrhlne--gviqagg--hghiggdvnrsvvrtvsam    1913
Qy    483  ---LYALKDNRTLQNLSDN-KSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKM        531
             |  |  |:                |:  |:|::
Qy   1914  eyfktplvsltaldnraglspatwnfqstyelid---ylldqryeyiwglyPtytew    1969
Qy    532  SI-------------MPSLDASPESHISLSLHFADAHQG------LLHG---KS       563
           |:                                          |:
Db   1970  svntlknldlgyyakpaptappmpkape--ldlrghtlesaegrkifgeykklqgeyeka  2027
Qy    564  ELEAQSV-AISHGRLVVADSEGKLFSAAI----PKQGDGNELKMKAMPQHALDEHFGH    616
Db   2028  kmavqaveaygeatrrvhdqlqryqkalggmdaetkevdg-------------iiqefaa  2075
Qy    617  DHQISGFFHDDHGQLNA---LVKNNFRQQ-HACPLGNDHQFHPGWNLTDALVIDNQLG--   670
           | |                :   |  |:|                 :|:  |:|:| ::
Db   2076  dlrtvyakqadqatidaetdkvagryksqidavrl--qaiqpg-rvtlakalsaalgad   2131
Qy    671  ---LHHTN-----------------PEPHEILDMGHLGSLALQEGKLHYFDQLTKGW     707
             :|:                                |:   :|:
Db   2132  wralghsqlmqrwkdfkagkrgaeiafypkeqtvlaag--agltlsngaih------nge  2183
Qy    708  TGAESDCKQ--LKKG---------LDGAAYLLKD-GEVKRLNINQSTSSIKHGTENVFSLPHV  758
           | :|:     :|: |         | |                  :|: :|:|: :|:
Db   2184  naaqnrgrpeglkigahsatsvsgsfdalrdvglekrldiddalaa-------vlvnphi  2236
Qy    759  -------------RNKPE----------PGDALQGLNKDDKAQAMAVI             783
                        | |:            | |
Db   2237  ftrigaaqtsladgaagpalarqarqapetdgmvdarglgsadalaslasldaaqglevs  2296
Qy    784  GVNKYLALTEKGDIRSFQIKPGTQOLERP---AQTLSREGISGELKD---IHVDHKQNLY   837
                |: :|:  |   |::|     |:|       |:  |:     |      |:|
Db   2297  g-rrnaqvadag------lagpsavaapavgaadvgvepvtgdqvdqpvvavgleqpva   2348
Db    838  ALTHEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQL   897
Qy   2349  tvrvappavalpr------------plfetrikfidqskfy------gsryff        2383
           || :|| :  :|            :|:  ||:|:            |  |::
Qy    898  KAGGWHAYAAPERGPLAVGTSGSQTVF------NRLMQGVKGKV-IPGSGLTVKLSAQTGG   951
Db   2384  eqigy----kpdraarvagdnyfdttlvreqvrralggyesrlpvrgvalvaklmdsagt  2439
             :|       |:|  :|   |::|        :|:   |  ||     |: |:
Qy    952  MTGAEGRKV---------SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLY  1004
Db   2440  vgkalglkvgvaptaqqlkgadrdfvwyvdtvidggkvlaprlylte--atrqgitdqy   2496
            |  |||:   : |      |: |  :|           |  |:| |      |:|
Qy   1005  EMQGALIK------QLDAHVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQ    1057
           |  |||:       ||                       | ||| |
Db   2497  agggaliasggdvtvntdghdv--ssvngliggrsvkvdagk-gkvvvadskgaggie-   2552
Qy   1058  SATRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPP  1117
Db   2553  -addevdvsgrdigi-eggklr-----------gkdvrlkadtvkvatsm           2589
            ::|                            |:|   |::|  :|
Qy   1118  SAESKLQSMLGHFVSAGVD-MSHQKGEIPLGRQRDPNDKTALT--KSRLILDTVTIGELH  1174
Db   2590  ryddk-----grlaargdaldaggqglhieakrletagatlkggkvkldvddvklggvy   2644
             :||      |: |    | :       ||  |  ::||  :|:|:|  :|::| ||:
Qy   1175  ELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTH--NKALEANY   1231
           |:  |  ||| |  |     |::|| |: |                      :|| |
Db   2645  eagss------yen---ksstplgslfailssttetnq                      2673
Qy   1232  DAVKAFINAFKKEHHGVNLTTRTV----LESQGSAELAKKLKNTLLSLDSGESMSFSR   1285
           || : |                    || |       |::| ::|:  |     || |
Db   2674  sa------hanhygtrieagtlegkmqnleieggsvda--ahtdlsvardarfkaaa    2722
Qy   1286  SYGGGVSTVFPTLSKKVPVPIPGAGITLDRAYNLSFSRTS-GGLNVSFGRDGGVSGNI   1344
           |:     |  |:                   :||           |:|  |  |:|::
Db   2723  dfahaehekdvrqlslgakv----gag----gyeagfslgsesgleahagr--gmtaga  2771
```

```
Db   1345  MVATGHDVMPYMT----                GKKTSAGNASDWLSAKHKISPDL---RIGAAV  1386
           |:                                          |:
Db   2772  evkvgyyrasheqsseteksyrnanlnfggsveagnvldigga------dinrryggaa  2825
Qy   1387  SGTLQGTLQNSLKFKLTEDELPGFIHGLT              HGTLTPAELLQKGI        1429
                |||    ||           |::                  ||||  |:|:|:
Db   2826  kgn-agt-eealrmrakkvestkyvseqtsqssgwsvevastasarsslltaatrlgdsv  2883
Qy   1430  EHQMKQGSKL----------             TFSVDTSANLDLRAGINLNEDGSKPNGVTA-----  1469
             ::|  |:|                        :|:|||| | |:    | :|:|
Db   2884  aqnvedgreirgelmaaqvaaeatqlvtadtaa-valsagisadfdsshsrstsqntqyl  2942
Qy   1470                          RVSAGLSASANLAAGSR--ERSTTSGQFGSTTSASN   1503
                                     |: ||:||  |  |::    ::|| |  | |||:|
Db   2943  ggnlsieategdatlvgakfgggdqvslkaaksvnlmaaestfesyseshnfhasadanl  3002
Qy   1504  NRPTFLNGVGAGANLTAALGVAHSSTHE-CKPVGIFPAFTSTNVSAA-LALDNRTSQSIS  1561
           |  |  |:|| | ||:|| |:|  ::     ::|  |:  |:::|:|  ||:|||||:|:
Db   3003  ganavggavglg-ltagmgtshgitnetgk------tyagtsvdaanvsid--agkdin   3052
Qy   1562  LELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAKD  1621
Db   3053  lsgsrvr-------------gkhvvldvegdinatskq                       3077
            ||| ||               ||:|:|||||:|:|::
Qy   1622  VVGDER-YE--------------AVRNLKKLVIRQQADSHSMELGSA--SHSTTYNNLSRI  1666
           | ||   |:              ||| ||||| ||||||||:|||    |||||||||||:
Db   3078  ---dernynssgggwdasagvaiqn-rtlv--------apvgsagfnfntehdn-srl    3122
Qy   1667  NNDGIVELL-------HKHFDAALPASSAKRLGEMNNDPALKDIIKQLQSTPFSSASVS   1719
           |||||    |       || |   | ||  ||| |:|||:|| ||:|||||:||||||:
Db   3123  tndgaagvvasdgltghvkgdanltgatiadlsgkgn---lk---------vdgavna   3168
Qy   1720  MELKDGLREQTEKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLG  1779
           ||||||||||||||||||||:|:  :||:|:||:|: ||    || ||:|| |:||||||
Db   3169  qnlkd-yrdk-----dggsgglnvg-------isstlaptvvafg----rvag        3206
Qy   1780  TSNSAAMSMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASALTDLKK          1831
           ||||||||||||| |:||||||||||||||||||||||| ||:|||||||||
Db   3207  edyqaeqratidvg--------qtkdparlqvggvvkgtlnqdaaqatvvqr          3250
```

RESULT 15
W04505
ID   W04505 standard; Protein; 1992 AA.
XX
AC   W04505;
XX
DT   25-JAN-1997 (first entry)
XX
DE   Moraxella 200 kDa outer membrane protein.
XX
KW   Outer membrane protein; OMP; immunogen; vaccine; otitis media;
KW   diagnosis.
XX
OS   Mycobacterium catarrhalis strain 4223.
XX
PN   WO9634960-A1.
XX
PD   07-NOV-1996.
XX
PF   29-APR-1996; 96WO-CA00264.
XX
PR   26-MAR-1996; 96US-0621944.
PR   01-MAY-1995; 95US-0431718.
PR   07-JUN-1995; 95US-0478370.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Chong P, Harkness RE, Klein MH, Loosmore SM, Sasaki K;
XX
DR   WPI; 1996-506162/50.
DR   N-PSDB; T38740.

```
XX
PT   Moraxella outer membrane protein - useful as immunogen in protective
PT   vaccine and for diagnosis
XX
PS   Claim 14; Fig 6; 109pp; English.
XX
CC   An approx. 200 kDa outer membrane protein (W04505) can be
CC   isolated from Moraxella catarrhalis otitis media strain 4223
CC   by electroelution, or expressed from a gene (see also T38740)
CC   obtd. from a strain 4223 genomic library. Natural or recombinant
CC   outer membrane protein is useful as an immunogen to protect
CC   against infection by Moraxella, esp. M. catarrhalis. It can
CC   also be used to detect antibodies, esp. for differential diagnosis
CC   between bacteria that cause similar symptoms, and also useful as
CC   a carrier for other antigens and used to raise antitumour
CC   antibodies for conjugation to therapeutic agents.
XX
SQ   Sequence  1992 AA;
```

```
Query Match        2.2%;   Score 204.5;  DB 17;   Length 1992;
Best Local Similarity  19.0%;   Pred. No. 0.00071;
Matches  318;   Conservative  221;  Mismatches  606;  Indels  531;  Gaps  79;
```

```
Qy   469  KDT-HSQLSRQ--------------ADG-KLYALKDNRTLQN------LSDNKSSEKLVDKIKS  510
               ::: ||     |  :|| :::||           |||:  |||: :: ::|
Db    15  kdtkhiaigeqnqprrsgtakadgdraiaigenanagggaiaigssnkt----------vng  67
```

```
Qy   511  YSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISL---SLHFADAHQGLLHGKSELEA  567
               :: || |   |          ::: |    :  |: :|: ||  : | || ::
Db    68  ssldkig------tdatgqesiaiggdvkasgdasiaigsddlhlldqhgnphpkgtlin  122
```

```
Qy   568  QSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDD  627
               :  | |     ||        ||  : :  :||  ||  |  |   ||
Db   123  dli--nghavlkei--------rsskdndvkyrrttas-----ghastavgamsya  163
```

```
Qy   628  HGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMG-  685
                 || |  |        |    |: || ::| :::|
Db   164  qghfs---naf-----------gtratataksaysavglaat-aeggstiaigsd  202
```

```
Qy   686  ---HLGSLALQEG---KLHYFDQLTKGWTGAESDCKQL------KKGLDGAAYLLKDGE  732
              |||  |: |          ||| |            |     ||  |  :| :|
Db   203  atsssLgaialgagtraqlqgsialgqgsvvtgsdnnsrpaytpntqaldpkfqatnntk  262
```

```
Qy   733  VKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDK---AQAMAVIGVNKYL  789
               |:| || |        :| |:        :|| :  ||  |||   |  |    || |
Db   263  agplsi--gsnsikrkiinvga--------------gvnktdavnvaqleavvkwaker  305
```

```
Qy   790  ALTEKGDIRSFQIKPGTQQ------LERPAQTLSREGISGELKD--IHVDHKQNLYALT  840
               ::||  : | :|| |     |     | |  |:|  || : ||   ::|  |||| |
Db   306  ritfggddnstdvkigldntltikggaetnaltdnnigvvkeadnsglkvklaktlnnlt  365
```

```
Qy   841  HEGEVFHQPREAWONGAESSSWHKLAL-------PQSESK------LKSLDMSHE  882
               :   |    |          | || |       |   |       |||  |  |
Db   366  evntlnatttvkvgsssttaellsdsltftqpntgsqstsktvygvngvkftnnaet  425
```

```
Qy   883  HKPIAT----------FEDG-----SQHQLKAG----------GWHAYAAPERGPLA  914
               :|  |          |   |     :|  ||          ||    ||
Db   426  taaigttritrdkigfardgdvdekqapyldkkqlkvgsvaitidngida-gnkkisnla  484
```

```
Qy   915  VGTSGSQTVFNRLMQGVKGKVIPGSGLTV------KLSAQTGGMT----------GAE  956
               :  :  |  |  ||    |:  :|  |       |:|  | ||           | |
Db   485  kgssandavtieqlkaakptlnagaqisvtpteisvdaksgnvtaptynigvkttelnsd  544
```

```
Qy   957  GRKVSSKFSERIRAYAFNPTMSTPRPI------------------------------  983
               ||  |  | |   || |  |   |
Db   545  g--tsdkfs--vkgsgtnnslvtaehlasylnevnrtadsalqsftvkeedddanaitv  600
```

```
Qy   984  ----KNAAYATQHGWQGREGLKPLYEMQGALIKQL-----------DAHNVRHNA  1023
               |    |||  |    | | ||  |           |
Db   601  akdttknagavsilklkgkngltvatkkdgtvtfglsqdsgltigkstlnndgltvkdtn  660
```

```
Qy  1024  PQPDLQS---KLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN  1080
```

```
Db   661  eqiqvgangikftnvngsnpgtgiantaritrdkigfags-------dgavdtnkpyl  711
                   :: |  |       |      | || : ||             :: |
Qy  1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ  1140
              | :::|      :    || |  ::   ||           | :|   |
Db   712  dqdklqvgnvkintginaggkait----glsptlpsi-------------adqssr  751
```

```
Qy  1141 KGEIPLGRQRDPNDKTALTKSRLJLDT----------------------VTIGEL  1173
              | |||  |  || |  :||| || |                        ||
Db   752 --nielgntiqdkdksnaasindilntgfnlknnnpidfvstydivdfangnattatvt  809
```

```
Qy  1174 HELADK-AKLVSD-------HKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHN  1224
              || || ||::: :|      ||||:| | || |:   |:  |: ::| ||    || |
Db   810  hdtanktskvvydvnvddtihltgtddnkklgvkttklnktsangn-----tatnfvn  864
```

```
Qy  1225 KALE--ANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLLSLDSGESM  1281
              |||   |:    |   |  :||  | ::|        | || | |||| ::|  |  || |
Db   865  ssdedalvnakdiaenlntlakeihttkgtadtalqt---ftvkkvdennaddanait  920
```

```
Qy  1282 SFSRSYGGGVSTVFVPTLSKKVPVVPIPGAGITLDRAYNLSFS-RTSGGLNV-SFGRDG  1338
              :::|  |  ||       :| | :    ||||  ||:   :::  ||: | |  ::| |
Db   921  vggknannqvntl--tlkge------nglniktdkngtvtfginttsglkagkstlndg  971
```

```
Qy  1339 GVS------GNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGT  1393
               ||:      | :  | |   :            |  |  |
Db   972  glsiknptgseqiqvgadgvkfa-------------kvnnngvvgaigdtt--  1010
```

```
Qy  1394 LQNSLKFLTEDELPGFIHGLTHGTLTPAE--LLQKGIEHQMKQGSKLTFSVDTSANLDL  1451
              |:|:  | |  | |  | ||::|  |:    |  :   | :  ||
Db  1011 ritrdei-gftg--tngsldkkphiskdgin--aggkkit-----ni  1048
```

```
Qy  1452 RAG-INLNE----DGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTSASNNRP  1506
              | |  ::|  |    |::|     |  | ::   ||   ::  |: ||: ||:  || |
Db  1049 qsgeiaqnshdavtggkiydlktelenkisstaktagnslhefsvadeqgnnftvsnpys  1108
```

```
Qy  1507 TF---------LNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALD  1553
                         |||  | :|  | :| ::  :   : ||  |   :|  || |  | |
Db  1109 sydtsktsdvitfagengittkvn----kgvvrvgidqtkgltpkltvgnnngkgivid  1164
```

```
Qy  1554 NRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSA--TTKMLAALKELDDAKPAEQL  1610
              : :::  ||   |||  ||   ::||  | ||: |    ||:  | |  ||:|   ||
Db  1165 sqngq-----------ntitglsntlanvtndkgsvrttegniikdedktraasiv  1210
```

```
Qy  1611 HILQQHFS----------AKDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLS  1664
              ::| :|            |:|  | |  :    :| |
Db  1211 dvlsagfnlqgngeavdfv--stydtv----------nfadgn---attakvtyddts  1253
```

```
Qy  1665 RIN---------NDGIVELLHKHF-----------------DAALPAS------  1686
              :::         :|  :| |||:                  |  ||
Db  1254 ktskvvydvnvddtievkdklgvktttltsgtgankfalsnqatgdalvkasdivah  1313
```

```
Qy  1687 ---------SAKRLGEMMNN----DPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKA  1733
                       :|::|  :||     | |: |  |    :   |::| ||  |::|| |
Db  1314 lntlsgdiqtakgasqannsagyvdadgnkviydstdnkyyqak---ndgtvdktkev  1368
```

```
Qy  1734 ILDGKVGREEV-GVLFQDRNNLRVKSVSVSQSV--SKSEGFNTP-ALLLGTSNSAAMS  1787
              | ||:  :|   ||:  |  |  |    |::||   ||  |    |:||| ::   |
Db  1369 akdklvaqaqtpdgtlaq---mnvksvinkeqvndankkgginednafvkglekaasdn  1424
```

```
Qy  1788 MERN------IGTIN----FKYGQDQNTPRR-------FTLEGGIAQANPQVASALTD  1828
              |:||      ||:::    ||  ||  |           | ||  | |  | || ||:
Db  1425 ktknaavtvgdlnavaqtpltfagdtgttakklgetlitkggqtdtn-----kltd  1475
```

Search completed: June 5, 2001, 18:16:36
Job time: 202 sec

```
                    GenCore version 4.5
        Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        June  5, 2001, 18:23:10 ; Search time 89.87 Seconds
                                          (without alignments)
                                          181.283 Million cell updates/sec

Title:         US-09-596-784-4
Perfect score: 715
Sequence:      1 MTSSQQRVERFLQYFSAGCK........IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters:      374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :

SPTREMBL_15:*
 1: sp_archea:*
 2: sp_bacteria:*
 3: sp_fungi:*
 4: sp_human:*
 5: sp_invertebrate:*
 6: sp_mammal:*
 7: sp_mhc:*
 8: sp_organelle:*
 9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_unclassified:*
13: sp_vertebrate:*
14: sp_virus:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	715	100.0	139	2	O54621		O54621 erwinia amy
2	455	63.6	124	2	Q9KH43		Q9kh43 erwinia her
3	273.5	38.3	129	2	O66102		O66102 pseudomonas
4	273.5	38.3	131	2	Q9JP37		Q9jp37 pseudomonas
5	80.5	11.3	368	10	Q40879		Q40879 pelargonium
6	80.5	11.3	766	4	Q9UKF4		Q9ukf4 homo sapien
7	80.5	11.3	787	4	Q9UKF3		Q9ukf3 homo sapien
8	80.5	11.3	820	4	Q9UKF5		Q9ukf5 homo sapien
9	80.5	11.3	820	4	Q9UHP1		Q9uhp1 homo sapien
10	78.5	11.0	830	2	Q9PE32		Q9pe32 xylella fas
11	77.5	10.8	775	2	Q9X7B0		Q9x7b0 mycobacteri
12	77	10.8	1166	4	Q9P2R3		Q9p2r3 homo sapien
13	75.5	10.6	402	5	O45260		O45260 caenorhabdi
14	75	10.5	2053	5	Q9XYM2		Q9xym2 drosophila
15	75	10.5	2053	5	Q9W5N4		Q9w5n4 drosophila
16	73	10.2	578	11	Q62693		Q62693 rattus norv
17	72.5	10.1	2183	14	Q98588		Q98588 subacute sc
18	72	10.1	377	10	Q9LT84		Q9lt84 arabidopsis
19	71.5	10.0	1591	4	Q92601		Q92601 homo sapien
20		71.5	10.0	5	Q9U347	2484	Q9u347 caenorhabdi
21		71.5	10.0	5	Q23187	2607	Q23187 caenorhabdi
22		71	9.9	11	Q9JIZ4	546	Q9jiz4 rattus norv
23		70.5	9.9	5	Q9NA71	1048	Q9na71 caenorhabdi
24		70.5	9.9	14	Q98334	2183	Q98334 measles vir
25		70.5	9.9	14	Q98589	2183	Q98589 measles vir
26		70.5	9.9	11	Q9WPY9	2183	Q9wpy9 measles vir
27		70	9.8	14	Q9JID5	594	Q9jid5 mus musculu
28		70	9.8	14	O56135	682	O56135 watermelon
29		70	9.8	3	Q9Y878	1860	Q9y878 coccidioide
30		69.5	9.7	5	Q9XTR0	291	Q9xtr0 caenorhabdi
31		69.5	9.7	8	Q37749	359	Q37749 cephalospor
32		69.5	9.7	2	Q55689	393	Q55689 synechocyst
33		69.5	9.7	14	Q83626	2183	Q83626 measles vir
34		69.5	9.7	14	Q9YZZ3	2183	Q9yzz3 measles vir
35		69.5	9.7	14	Q9YZY9	2183	Q9yzy9 measles vir
36		69.5	9.7	14	Q9WPZ1	2183	Q9wpz1 measles vir
37		69.5	9.7	14	Q9WPY8	2183	Q9wpy8 measles vir
38		69.5	9.7	14	Q9WPY7	2183	Q9wpy7 measles vir
39		69.5	9.7	14	Q9WPY6	2183	Q9wpy6 measles vir
40		69.5	9.7	14	Q9WMB3	2183	Q9wmb3 measles vir
41		69.5	9.7	14	Q9WPZ0	2183	Q9wpz0 measles vir
42		69.5	9.7	14	Q9W984	2183	Q9w984 measles vir
43		69.5	9.7	14	Q9IFK4	2183	Q9ifk4 measles vir
44		69.5	9.7	14	Q9IC35	2183	Q9ic35 measles vir
45		69.5	9.7	6	O97592	3680	O97592 canis famil

ALIGNMENTS

```
RESULT    1
O54621
ID   O54621      PRELIMINARY;    PRT;   139 AA.
AC   O54621;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   DSPF.
GN   DSPF OR DSPB.
OS   Erwinia amylovora.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Erwinia.
OX   NCBI_TaxID=552;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EA321;
RX   MEDLINE=98115919; PubMed=9448330;
RA   Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
RA   Conlin A.K., Collmer A., Beer S.V.;
RT   "Homology and functional similarity of an hrp-linked pathogenicity
RT   locus, dspEF, of Erwinia amylovora and the avirulence locus avrE of
RT   Pseudomonas syringae pathovar tomato.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EA321;
RA   Bogdanove A.J., Kim J.F., Wei Z.-M., Kolchinsky P., Beer S.V.;
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CFBP1430;
RX   MEDLINE=98086111; PubMed=9426142;
RA   Gaudriault S., Malandrin L., Paulin J.P., Barny M.A.;
RT   "DspA, an essential pathogenicity factor of Erwinia amylovora showing
RT   homology with AvrE of Pseudomonas syringae, is secreted via the Hrp
RT   secretion pathway in a DspB-dependent way.";
RL   Mol. Microbiol. 26:1057-1069(1997).
DR   EMBL; U97504; AAC04851.1; -.
DR   EMBL; Y13831; CAA74157.1; -.
SQ   SEQUENCE   139 AA;  15723 MW;  60F23F88B1BFAE93 CRC64;
```

Query Match   100.0%;   Score 715; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 4.3e-68;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60

Qy   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120

Qy 121 EHAAEVREYIAQLDESSAA 139
        |||||||||||||||||||
Db 121 EHAAEVREYIAQLDESSAA 139

RESULT 2
Q9KH43
ID   Q9KH43   PRELIMINARY;   PRT;   124 AA.
AC   Q9KH43;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   DSPF.
GN   DSPF.
OS   Erwinia herbicola pv. gypsophilae.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Erwinia.
OX   NCBI_TaxID=48984;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mor H., Manulis S., Zuck M., Nizan R., Coplin D.L., Barash I.;
RT   "Genetic Organization of the hrp Gene Cluster and dspEF Operon in
RT   Erwinia herbicola pv. gypsophilae.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF271717; AAF76344.1; -.
SQ   SEQUENCE   124 AA; 13873 MW; 9F0266FCB78CCEB0 CRC64;

Query Match   63.6%;   Score 455; DB 2; Length 124;
Best Local Similarity   67.5%; Pred. No. 1.1e-40;
Matches 81; Conservative 18; Mismatches 21; Indels 0; Gaps 0;

Qy   1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
        |||||||| : |||||: :| |||||| || : :: :|||||||| | |||||
Db   1 MTPSQQRVDAFLQHFEAASCKTPLHLKEGVCPMYTDQGEEAAVLEVPEHSDCLLLHCELAS 60

Qy   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
        :   :: ||:|LLQLNFEMAAMRGCWLALDE ||:| |||||||| :: :|| |||||
Db   61 TAPHSALNFYALLLQLNFEMAAMRGCWLALDEQHKVRLCFQQSLAQMDETNFSQLLSGFI 120

RESULT 3
O66102
ID   O66102   PRELIMINARY;   PRT;   129 AA.
AC   O66102;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   AVIRULENCE PROTEIN.
GN   AVRF.
OS   Pseudomonas syringae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=317;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=9811591; PubMed=9448330;
RA   Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
RA   Conlin A.K., Collmer A., Beer S.V.;
RT   "Homology and functional similarity of an hrp-linked pathogenicity

RT   locus, dspEF, of Erwinia amylovora and the avirulence locus avrE of
RT   Pseudomonas syringae pathovar tomato.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).
DR   EMBL; U97505; AAC06135.1; -.
SQ   SEQUENCE   129 AA; 14339 MW; 19ABBBD84968ECE9 CRC64;

Query Match   38.3%;   Score 273.5; DB 2; Length 129;
Best Local Similarity   43.1%; Pred. No. 1.7e-21;
Matches 56; Conservative 27; Mismatches 44; Indels 3; Gaps 0;

Qy   1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
        | ||: | : : ||||| |: : |:|:::: :|| ::||:|: ||||
Db   1 MKTSQPDFARFINSLGAQLGTLTLQNGVCALYDGQNNEAAIIELPEHSEMVIFHCRIGR 60

Qy   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
        | : | ||| ::| :|| | : |||| | :|||| :| |
Db   61 C-PERAPDLL-RLLSLNFDVARLHGCWFAVDQ-GDVRLCAQRELASLDEPAFCDVTRGFI 117

Qy 121 EHAAEVREYI 130
        ::: ::
Db 118 SQAREARAFL 127

RESULT 4
Q9JP37
ID   Q9JP37   PRELIMINARY;   PRT;   131 AA.
AC   Q9JP37;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HYPOTHETICAL 14.6 KDA PROTEIN.
OS   Pseudomonas syringae (pv. tomato).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=9842276; PubMed=9748456;
RA   Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA   Collmer A.;
RT   "The Pseudomonas syringae pv. tomato Hrpw protein has domains similar
RT   to harpins and pectate lyases and can elicit the plant hypersensitive
RT   response and bind to pectate.";
RL   J. Bacteriol. 180:5211-5217(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=20243785; PubMed=10781092;
RA   Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
RA   Petnicki-Ocwieja T., van Dijk K., Collmer A.;
RT   "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT   mosaic structure composed of a cluster of type III secretion genes
RT   bounded by exchangeable effector and conserved effector loci that
RT   contribute to parasitic fitness and pathogenicity in plants.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
DR   EMBL; AF232006; AAF71500.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   131 AA; 14571 MW; F2A34660EF98EC98 CRC64;

Query Match   38.3%;   Score 273.5; DB 2; Length 131;
Best Local Similarity   43.1%; Pred. No. 1.8e-21;
Matches 56; Conservative 27; Mismatches 44; Indels 3; Gaps 0;

Qy   1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
        | ||: | : : ||||| |: : |:|:::: :|| ::||:|: ||||
Db   3 MKTSQPDFARFINSLGAQLGTLTLQNGVCALYDGQNNEAAIIELPEHSEMVIFHCRIGR 62

Qy   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
        | : | ||| ::| :|| | : |||| | :|||| :| |
Db   63 C-PERAPDLL-RLLSLNFDVARLHGCWFAVDQ-GDVRLCAQRELASLDEPAFCDVTRGFI 119

```
Qy  121 EHAAEVREYI 130
            | || :: |
Db  120 SQAREARAFL 129

RESULT  5
Q40879
ID  Q40879          PRELIMINARY;     PRT;   368 AA.
AC  Q40879;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE  MYRISTYL-ACP DESATURASE.
OS  Pelargonium hortorum.
OC  Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC  Magnoliophyta; eudicotyledons; core eudicots; Rosidae; Geraniales;
OC  Geraniaceae; Pelargonium.
OX  NCBI_TaxID=4031;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=LINE: 88-51-10; TISSUE=TRICHOME GLANDS;
RA  Schultz D.J., Cahoon E.B., Shanklin J., Craig R., Cox-Foster D.L.,
RA  Mumma R.O., Medford J.I.;
RL  Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; U40344; AAC49421.1; -.
DR  HSSP; P22337; 1AFR.
DR  MENDEL; 9742; Pelho;1039;9742.
DR  INTERPRO; IPR001225; -.
DR  PFAM; PF00487; FA_desaturase; 1.
SQ  SEQUENCE 368 AA;  42145 MW;  8A778B920883D99E CRC64;

Query Match          11.3%;   Score 80.5;  DB 10;  Length 368;
Best Local Similarity  20.1%;  Pred. No. 1.4;
Matches  37;  Conservative  26;  Mismatches  54;  Indels  67;  Gaps      8;

Qy    6 QRVERFLQYFSA----GC-KTPIHLKDGVCALYNEQDEEAAVLE------VPQHSDSL 52
          :  | | |     |  | | ||  | :  ||    ||      ||||||||
Db  169 RQVEKTIQYLIALGQDIGTEKNPYHL-----FIYTSFQERATFISHANTAKLAQQHGDKQ 223

Qy   53 LLH-CRIIEADPQTSITLYSMLLQLNFEM----- 80
          | |  |  | |  |  | | | | ||
Db  224 LAQICGTIAADEKRHETAYTRIVDKLFELDPDETMSCLAHMKRKITMPAHLMRDGRDPH 283

Qy   81 ------:-AAMRGCWLALDELHNVRLCFQQSLEH-LDEASFSDIVSGFIEHAAEVREYIA 131
            |  ::  ||  | | : |  : :| | | | :: : ::
Db  284 LFQHFSVVASRTGVYTVWDYIN--------ILEHFVEKWNIEKITAGLSDKGREAQDYVC 335

Qy  132 QLDE 135
          | |
Db  336 KLGE 339

RESULT  6
Q9UKF4
ID  Q9UKF4          PRELIMINARY;     PRT;   766 AA.
AC  Q9UKF4;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE  METALLAPROTEINASE-DISINTEGRIN BETA.
GN  ADAM29.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Cerretti D.P., DuBose R.F., Black R.A., Nelson N.;
RT  "Isolation of Two Novel Metalloproteinase-Disintegrin (ADAM) cDNAs
RT  That Show Testis-Specific Gene Expression.";
RL  Biochem. Biophys. Res. Commun. 263:810-815(1999).
DR  EMBL; AF171930; AAF03778.1; -.
DR  HSSP; P18619; 1FVL.
DR  INTERPRO; IPR001590; -.
DR  INTERPRO; IPR001762; -.
DR  INTERPRO; IPR002870; -.
DR  PFAM; PF00200; disintegrin; 1.
DR  PFAM; PF01421; Reprolysin; 1.
DR  PFAM; PF01562; Pep_M12B_propep; 1.
DR  PRINTS; PR00289; DISINTEGRIN.
KW  Integrin.
SQ  SEQUENCE 766 AA;  86982 MW;  ACA4BEDFA700ABA5 CRC64;

Query Match          11.3%;   Score 80.5;  DB 4;  Length 766;
Best Local Similarity  22.3%;  Pred. No. 3.4;
Matches  31;  Conservative  28;  Mismatches  47;  Indels  33;  Gaps      6;

Qy   18 GCKTPIHLK------DGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITL-- 69
          | ::|:|                                              :|
Db   60 GQKHIHIKVKKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYH-GYVEGDPESLVSLST 118

Qy   70 ----YSMLLQLNFEMAAMRGCWLALDELHNVR-LCFQQSLEHL-----DEASFSDIVSG 118
            :||:|                                      |||
Db  119 CFGGFQGIIQIN--------DFAYEIKPLAFSTTFEHLVYKMDSEEKQFSTMRSG 165

Qy  119 FIEHAAEVREYIAQLDESS 137
          | :::      :  :| |
Db  166 FMQNEITCRMEFEEIDNST 184

RESULT  7
Q9UKF3
ID  Q9UKF3          PRELIMINARY;     PRT;   787 AA.
AC  Q9UKF3;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE  METALLAPROTEINASE-DISINTEGRIN GAMA.
GN  ADAM29.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Cerretti D.P., DuBose R.F., Black R.A., Nelson N.;
RT  "Isolation of Two Novel Metalloproteinase-Disintegrin (ADAM) cDNAs
RT  That Show Testis-Specific Gene Expression.";
RL  Biochem. Biophys. Res. Commun. 263:810-815(1999).
DR  EMBL; AF171931; AAF03779.1; -.
DR  HSSP; P18619; 1FVL.
DR  INTERPRO; IPR001590; -.
DR  INTERPRO; IPR001762; -.
DR  INTERPRO; IPR002870; -.
DR  PFAM; PF00200; disintegrin; 1.
DR  PFAM; PF01421; Reprolysin; 1.
DR  PFAM; PF01562; Pep_M12B_propep; 1.
DR  PRINTS; PR00289; DISINTEGRIN.
KW  Integrin.
SQ  SEQUENCE 787 AA;  89459 MW;  2B0E5E7C496B8C9D CRC64;

Query Match          11.3%;   Score 80.5;  DB 4;  Length 787;
Best Local Similarity  22.3%;  Pred. No. 3.5;
Matches  31;  Conservative  28;  Mismatches  47;  Indels  33;  Gaps      6;

Qy   18 GCKTPIHLK------DGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITL-- 69
          | ::|:|                                              :|
Db   60 GQKHIHIKVKKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYH-GYVEGDPESLVSLST 118

Qy   70 ----YSMLLQLNFEMAAMRGCWLALDELHNVR-LCFQQSLEHL-----DEASFSDIVSG 118
            :||:|                                      |||
Db  119 CFGGFQGIIQIN--------DFAYEIKPLAFSTTFEHLVYKMDSEEKQFSTMRSG 165
```

Qy   119 FIEHAAEVREYIAQLDESS 137
            |::| | :|| |
Db   166 FMQNEITCRMEFEEIDNST 184

RESULT 8
Q9UKF5
ID   Q9UKF5   PRELIMINARY;   PRT;   820 AA.
AC   Q9UKF5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   METALLAPROTEINASE-DISINTEGRIN.
GN   ADAM29.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cerretti D.P., DuBose R.F., Black R.A., Nelson N.;
RT   "Isolation of Two Novel Metalloproteinase-Disintegrin (ADAM) cDNAs
RT   That Show Testis-Specific Gene Expression.";
RL   Biochem. Biophys. Res. Commun. 263:810-815(1999).
DR   EMBL; AF171929; AAF03777.1; -.
DR   HSSP; P18619; 1FVL.
DR   INTERPRO; IPR001590; -.
DR   INTERPRO; IPR001762; -.
DR   INTERPRO; IPR002870; -.
DR   PFAM; PF00200; disintegrin; 1.
DR   PFAM; PF01421; Reprolysin; 1.
DR   PFAM; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
KW   Integrin.
SQ   SEQUENCE   820 AA;   92781 MW;   B054E9FAC05E4EF5 CRC64;

Query Match   11.3%; Score 80.5; DB 4; Length 820;
Best Local Similarity   22.3%; Pred. No. 3.7;
Matches   31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

Qy   18 GCKTPIHLK------DGVCALYNEQDEEAAVLEVPQHSDSLLHCRIIEADPQTSITL-- 69
            | ||| |:|| :: : |:| : :|| ||::: :||:
Db   60 GQKHIHIKVKKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYYH-GYVEGDPESLVSLST 118

Qy   70 ---YSMLLQLNFEMAAMRGCWLALDELHNVR-LCFQQSLEHL------DEASFSDIVSG 118
               ::||:| | :: ||| | || || | || ::|
Db   119 CFGGFQGILQIN--------------------DFAYEIKPLAFSTTFEHLVYKMDSEEKQFSTMRSG 165

Qy   119 FIEHAAEVREYIAQLDESS 137
            |::| | :|| |
Db   166 FMQNEITCRMEFEEIDNST 184

RESULT 9
Q9UHP1
ID   Q9UHP1   PRELIMINARY;   PRT;   820 AA.
AC   Q9UHP1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   DISINTEGRIN AND METALLOPROTEINASE DOMAIN 22.
GN   ADAM22.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Xu R., Cai J., Xu T., Zhou W., Zhao S., Li C.;
RT   "Cloning of a novel cDNA sequence of the family of ADAM.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.

DR   EMBL; AF134708; AAF22163.1; -.
DR   HSSP; P18619; 1FVL.
DR   INTERPRO; IPR001590; -.
DR   INTERPRO; IPR001762; -.
DR   INTERPRO; IPR002870; -.
DR   PFAM; PF00200; disintegrin; 1.
DR   PFAM; PF01421; Reprolysin; 1.
DR   PFAM; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
KW   Integrin.
SQ   SEQUENCE   820 AA;   92900 MW;   EE6B4CB9EBDA2275 CRC64;

Query Match   11.3%; Score 80.5; DB 4; Length 820;
Best Local Similarity   22.3%; Pred. No. 3.7;
Matches   31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

Qy   18 GCKTPIHLK------DGVCALYNEQDEEAAVLEVPQHSDSLLHCRIIEADPQTSITL-- 69
            | ||| |:|| :: : |:| : :|| ||::: :||:
Db   60 GQKHIHIKVKKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYYH-GYVEGDPESLVSLST 118

Qy   70 ---YSMLLQLNFEMAAMRGCWLALDELHNVR-LCFQQSLEHL------DEASFSDIVSG 118
               ::||:| | :: ||| | || || | || ::|
Db   119 CFGGFQGILQIN--------------------DFAYEIKPLAFSTTFEHLVYKMDSEEKQFSTMRSG 165

Qy   119 FIEHAAEVREYIAQLDESS 137
            |::| | :|| |
Db   166 FMQNEITCRMEFEEIDNST 184

RESULT 10
Q9PE32
ID   Q9PE32   PRELIMINARY;   PRT;   830 AA.
AC   Q9PE32;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE ALPHA CHAIN.
GN   XF1196.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

```
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-157(2000).
DR   EMBL; AE003954; AAF84006.1; -.
DR   INTERPRO; IPR000788; -.
DR   PFAM; PF00317; ribonucleo_red; 1.
DR   PRINTS; PR01183; RIBORDTASEM1.
DR   PROSITE; PS00089; RIBORED_LARGE; 1.
SQ   SEQUENCE   830 AA;  94229 MW;  12207407246E28BC CRC64;

Query Match             11.0%;  Score 78.5;  DB 2;  Length 830;
Best Local Similarity   25.6%;  Pred. No. 6.1;
Matches   33;  Conservative   21;  Mismatches   46;  Indels   29;  Gaps    6;

Qy     17 AGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRII-EADPQTSITLYSMLLQ 75
           :  | :||:||   ::|:|:  |::||    :  |::|
Db     22 AGSVPPVALS----VFHPDEPED---EVPIPRQTAMMHTRAVEEANVSTWITKEAGNRR 73

Qy     76 LNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEV-------- 126
           :  |:|.||   :|:|:|     ||:||
Db     74 MPFDCARLE---CAIDTIH------QEFPQLDVAEYKRAVFGFVERKDSVNADDLVDLL 123

Qy    127 -REYIAQLD 134
           || |::|
Db    124 IREAEARVD 132

RESULT 11
Q9X7B0
ID   Q9X7B0.
AC   Q9X7B0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   PUTATIVE GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE.
GN   PLSB.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93188700; PubMed=8446027;
RA   Seeger K.J., Harris D.;
RT   "Use of an ordered cosmid library to deduce the genomic organization
RT   of Mycobacterium leprae.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93188700; PubMed=8446027;
RA   Eiglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT   "Use of an ordered cosmid library to deduce the genomic organization
RT   of Mycobacterium leprae.";
RL   Mol. Microbiol. 7:197-206(1993).
DR   EMBL; AL049913; CAB43153.1; -.
DR   INTERPRO; IPR002123; -.
DR   PFAM; PF01553; Acyltransferase; 1.
KW   Transferase; Acyltransferase.
SQ   SEQUENCE   775 AA;  87363 MW;  03DD77C778293CDF CRC64;

Query Match             10.8%;  Score 77.5;  DB 2;  Length 775;
Best Local Similarity   25.8%;  Pred. No. 7.1;
Matches   34;  Conservative   16;  Mismatches   33;  Indels   49;  Gaps    5;

Qy     11 FLQYFSAGCKTPIHLKDGVCALYN---------------------------EQ 36
           |   ||| ||| ||:|:
Db    397 YAAYARGGEKTP---EGVAWLYSFIKAQGERNYGKIVRFPEAVSMRQYLGAPHGALVQ 452
```

```
Qy     37 DEEAAVLEVPQHSDSLLLHCRIIEADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNV 96
           |:|:||:||| ||:||||:||::||  ||:|  |::|::||
Db    453 DQDAKRLALQKMSFEVA-WRILCATPVTATALVSALL-----LTTRGVALTLDQLHHT 504

Qy     97 RLCFQQSLEHLD 108
           |:||:|:|:
Db    505 --LQESLDYLE 513

RESULT 12
Q9P2R3
ID   Q9P2R3.                   PRELIMINARY;   PRT;  1166 AA.
AC   Q9P2R3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   ANKHZN.
GN   ANKHZN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kuriyama H., Maruyama H., Ishii N., Ito K., Kuwano R.;
RT   "Homo sapiens ANKHZN gene.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037360; BAA90300.1; -.
SQ   SEQUENCE   1166 AA;  127902 MW;  B595E8DE88D77B0A CRC64;

Query Match             10.8%;  Score 77;  DB 4;  Length 1166;
Best Local Similarity   24.8%;  Pred. No. 13;
Matches   38;  Conservative   27;  Mismatches   38;  Indels   50;  Gaps   10;

Qy     11 FLQYFSAGC-KTPIHLKDGVCALYNEQDEEAAVL-EVPQHSDSLL------------ 53
           ||:| |:  |||::||:|||   ::||  :|:|::||
Db    312 FVNAATLGCQETPLHL---VALYSSRNHSADVMSEMAQIAEALLQAGANPNMQDSKGRT 367

Qy     54 -LHCRIIEADPQTSITLYSMLL--QLNFEMAAMRG---CWLALDELHNVRLCFQQSLEH 106
           :|:||  ||| ::|  |||   ::|   |::| |||
Db    368 PLHVSIMAGDEY---VFSHLLQCKQLDLELKDHEGSTALWLA---VQHITVSSDQSVNP 420

Qy    107 LDEASFSDIVSGFIEHAAEVREYIAQLDESSAA 139
           :  :|:|    |:|:|
Db    421 FEDV--PVVNG---------TSFDENSFA 438

RESULT 13
O45260
ID   O45260.                   PRELIMINARY;   PRT;   402 AA.
AC   O45260;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   C18D4.2 PROTEIN.
GN   C18D4.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mortimore B.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
```

RA Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RT Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL; Z81474; CAB03900.1; -.
DR INTERPRO; IPR002900; -.
DR PFAM; PF01827; DUF38; 1.
SQ SEQUENCE 402 AA; 47760 MW; F6A918FC72EB96CE CRC64;

Query Match   10.6%;   Score 75.5; DB 5; Length 402;
Best Local Similarity 24.3%; Pred. No. 5.4;
Matches   33; Conservative 23; Mismatches 35; Indels 45; Gaps 8;

Qy   5 QQRVERFLQYF----SAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSDSLLL---HC 56
           :||:|:||:| || | || : || :|| | ||||
Db 202 EQRIDLPVKYFEDILSSVGHTIPVE-----ALYMQACNQNTVLRVLRHLDSRLLKTISI 255

Qy   57 RIEADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIV 116
         ::|:||| | ||||| -EIC-----ELDQWKFAENI 289
Db 256 KIVE-DDQDKPCNWVLPLEIN----------EIC-----ELDQWKFAENI 289

Qy 117 SGFIEH--AAEVREY 129
         :|| :||
Db 290 E--VEHLIDTDDVRHF 303

RESULT 14
Q9XYM2
ID Q9XYM2   PRELIMINARY;   PRT;   2053 AA.
AC Q9XYM2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE ADHERIN NIPPED-B.
GN NIPPED-B.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Rollins R.A., Morcillo P., Dorsett D.;
RT "Nipped-B, a Drosophila homolog of chromosomal adherins, participates
RT in activation by remote enhancers in the cut and ultrabithorax
RT genes.";
RL Genetics 0:0-0(1999).
DR EMBL; AF114160; AAD26161.1; -.
SQ SEQUENCE 2053 AA; 233927 MW; AC3E6C424A2D8F41 CRC64;

Query Match   10.5%;   Score 75; DB 5; Length 2053;
Best Local Similarity 20.5%; Pred. No. 41;
Matches   32; Conservative 28; Mismatches 28; Indels 50; Gaps 7;

Qy   12 LQYFSAGCKTPIHL--KDGV-------------CALYNEQDEEAAVLEVPQ-------- 47
           |:|| || | | ||| |: |||: || || :||
Db 1490 MRYFD-FKSPIALGETNDGLPVSICEDVFHCLMFFCRCTNQEIRKQALISLGSFCVLND 1547

Qy   48 ---HSDSLLHCRIIEADPQTSITLYSMLLQLNFEMAAMRGCWLALDE----LHNVRLC 99
            :|| ||:   :||| :| ::||| |: || :|| :| ||
Db 1548 GYLTRSELKNLYCEI-----------LSSIANDAGFKIICMRNIWIYLTESEMFMHNKEKE 1597

Qy 100 FQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQLDE 135
          : || |:||| ::| |: :||
Db 1598 WEKQSKHEDLKEMNDVSSGMASRIIQL-YLEEILE 1631

RESULT 15
Q9W5N4
ID Q9W5N4   PRELIMINARY;   PRT;   2053 AA.
AC Q9W5N4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE NIPPED-B PROTEIN.
GN NIPPED-B.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003040; AAF45353.1; -.
DR FLYBASE; FBgn0026401; Nipped-B.
SQ SEQUENCE 2053 AA; 233880 MW; 47F03lE13ACF9BDB CRC64;

Query Match   10.5%;   Score 75; DB 5; Length 2053;
Best Local Similarity 20.5%; Pred. No. 41;
Matches   32; Conservative 28; Mismatches 28; Indels 50; Gaps 7;

Qy   12 LQYFSAGCKTPIHL--KDGV-------------CALYNEQDEEAAVLEVPQ-------- 47
           |:|| || | | ||| |: |||: || || :||
Db 1490 MRYFD-FKSPIALGETNDGLPVSICEDVFHCLMFFCRCTNQEIRKQALISLGSFCVLND 1547

Qy   48 ----HSDSLLHCRIIEADPQTSITLYSMLLQLNFEMAAMRGCWLALDE----LHNVRLC 99
             :|| ||:   :||| :| ::||| |: || :|| :| ||
Db 1548 GYLTRSELKNLYCEI-------------LSSIANDAGFKIICMRNIWIYLTESEMFMHNKEKE 1597

Qy   100 FQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQLDE 135
          ::: :|| :|: || ::: || ::|| :::|:|
Db   1598 WEKQSKHEDLKEMNDVSSGMASRIIQL--YLEEILE 1631

Search completed: June 5, 2001, 18:23:11
Job time: 392 sec

```
                    GenCore version 4.5
          Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    June 5, 2001, 18:23:08 ; Search time 28.9 Seconds
                                    (without alignments)
                                    164.758 Million cell updates/sec

Title:          US-09-596-784-4
Perfect score:  715
Sequence:       1 MTSSQQRVERFLQYFSAGCK.........IEHAAEVREYIAQLDESSAA 139

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters:   93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	79	11.0	789	1	PLS2_MYCTU	O53207 mycobacteri
2	75.5	10.6	966	1	SSN6_YEAST	P14922 saccharomyc
3	73.5	10.3	320	1	DPA4_BPR69	O64301 bacteriopha
4	73.5	10.3	976	1	EPA1_HUMAN	P21709 homo sapien
5	71.5	10.0	920	1	OGT1_HUMAN	O15294 homo sapien
6	71.5	10.0	1036	1	OGT1_RAT	P56558 rattus norv
7	71		816	1	NEL2_MOUSE	Q61220 mus musculu
8	70.5	9.9	2203	1	POLG_EC09B	Q66577 e genome po
9	70	9.8	946	1	AMPN_PLUXY	P91887 plutella xy
10	69.5	9.7	2183	1	RRPL_MEASA	P35975 measles vir
11	69.5	9.7	2183	1	RRPL_MEASE	P12576 measles vir
12	69	9.7	451	1	SUN_HAEIN	P44788 haemophilus
13	68.5	9.6	595	1	MERL_HUMAN	P35240 homo sapien
14	68.5	9.6	596	1	MERL_MOUSE	P46662 mus musculu
15	68	9.5	478	1	ARDE_CHLMU	P56961 chlamydia m
16	68	9.5	1141	1	HEM2_CAEEL	P55163 caenorhabdi
17	67.5	9.4	1527	1	MRP3_HUMAN	O15438 homo sapien
18	67.5	9.4	532	1	UD14_RABIT	Q28612 oryctolagus
19	67	9.4	578	1	YQD6_CAEEL	Q09266 caenorhabdi
20	67	9.3	283	1	DMPD_PSESP	P19076 pseudomonas
21	66.5	9.3	412	1	NEMO_MOUSE	O88522 mus musculu
22	66.5	9.3	925	1	DBL_HUMAN	P10911 homo sapien
23	66.5	9.3	1174	1	YJI1_YEAST	P47108 saccharomyc
24	66	9.2	617	1	FILS_RAT	Q02435 rattus norv
25	66	9.2	644	1	RNR_HELPY	P56123 helicobacte
26	66	9.2	3828	1	TRX_DROVI	Q24742 drosophila
27	65.5	9.2	334	1	OTCC_RHIET	O31018 rhizobium e
28	65.5	9.2	416	1	SYH_METJA	Q58406 methanococc
29	65.5	9.2	477	1	HRPB_BURSO	P31778 burkholderi
30	65.5	9.2	947	1	PODK_MAIZE	P11155 zea mays (m
31	65.5	9.2	5035	1	RYNR_PIG	P16960 sus scrofa
32	65	9.1	143	1	Y199_MYCPN	P75446 mycoplasma
33	65	9.1	244	1	DCOP_BACCL	P46535 bacillus ca
34	65	9.1	472	1	DAX1_RAT	P70503 rattus norv
35	65	9.1	750	1	DPOL_HPBVZ	P03155 hepatitis b
36	65	9.1	1022	1	PIB4_HUMAN	Q15147 homo sapien
37	65	9.1	1023	1	PIB4_BOVIN	Q07722 bos taurus
38	65	9.1	1175	1	PIB4_RAT	Q9qw07 rattus norv
39	65	9.1	1756	1	TRI2_ECOLI	P22706 escherichia
40	65	9.1	2649	1	BPA1_HUMAN	Q33001 homo sapien
41	65	9.1	5032	1	RYNR_HUMAN	P21817 homo sapien
42	64.5	9.0	370	1	RF1_MYCCA	P71496 mycoplasma
43	64.5	9.0	533	1	YA83_MYCPN	P75610 mycoplasma
44	64.5	9.0	587	1	RECN_MYCLE	Q49896 mycobacteri
45	64.5	9.0	818	1	SAP4_YEAST	P53036 saccharomyc

ALIGNMENTS

```
RESULT 1
PLS2_MYCTU
ID   PLS2_MYCTU     STANDARD;      PRT;   789 AA.
AC   O53207;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PUTATIVE ACYLTRANSFERASE PLSB2.
GN   PLSB2 OR RV2482C OR MTV008.38C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE GPAT / DAPAT FAMILY.
CC
CC   -!- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
DR   EMBL; AL021246; CAA16059.1; -.
DR   TubercuList; Rv2482c; -.
DR   InterPro; IPR002123; -.
DR   Pfam; PF01553; Acyltransferase; 1.
KW   Hypothetical protein; Transferase; Acyltransferase.
SQ   SEQUENCE   789 AA;  88314 MW;  B78D7D8F8296EA22 CRC64;

Query Match             11.0%;  Score 79;  DB 1;  Length 789;
Best Local Similarity   24.3%;  Pred. No. 2.3;
Matches   34;  Conservative   23;  Mismatches   51;  Indels   32;  Gaps    4;

Qy      23 IHLKDGVCALYNEQDERAAVLEVPQHSDSLLHCRIIEADPQTSITLYSMLLQLNFEMAA 82
              :  . :.:  ..: ||||||||||      :  ||  || :| ::|
Db     445 VSMRQYIGAPHGELTQDPAAKRIALQKMSFEVAWRILQATPVTATGLVSALL------LT 498

Qy      83 MRGCWLALDELHNVRLCFQQSLEHLD----------------------EASFSDIVSGF---- 119
           ||  |  |:|||::  :|:|:|:            |:  .:|:|
```

```
Db    499 TRGTALTLDQLHHT---LQDSLDYLERKQSPVSTSALRLRSREGVRAAADALSNGHPVTR 555

Qy    120 IEHAAEVREYIAQLDESSAA 139
          :: .|| |||| || :||
Db    556 VDSGREPVWYIAPDDEHAAA 575


RESULT  2
SSN6_YEAST
ID   SSN6_YEAST    STANDARD;    PRT;   966 AA.
AC   P14922;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   GLUCOSE REPRESSION MEDIATOR PROTEIN.
GN   SSN6 OR CYC8 OR YBR112C OR YBR0908.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89211964; PubMed=2854095;
RA   Trumbly R.J.;
RT   "Cloning and characterization of the CYC8 gene mediating glucose
RT   repression in yeast.";
RL   Gene 73:97-111(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88065502; PubMed=3316983;
RA   Schultz J., Carlson M.;
RT   "Molecular analysis of SSN6, a gene functionally related to the SNF1
RT   protein kinase of Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 7:3637-3645(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=92327848; PubMed=1626431;
RA   Mannhaupt G., Stucka R., Ehnle S., Vetter I., Feldmann H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
RN   [4]
RP   TPR REPEATS.
RX   MEDLINE=90124639; PubMed=2404612;
RA   Sikorski R.S., Boguski M.S., Goebl M., Hieter P.A.;
RT   "A repeating amino acid motif in CDC23 defines a family of proteins
RT   and a new relationship among genes required for mitosis and RNA
RT   synthesis.";
RL   Cell 60:307-317(1990).
CC   -!- FUNCTION: IT IS INVOLVED IN REPRESSION BY A1-ALPHA2 AND ALPHA2 AND
CC       IN OTHER SYSTEMS AS A GENERAL REPRESSOR OF TRANSCRIPTION. THIS
CC       PROTEIN HAS NO OBVIOUS DNA-BINDING DOMAINS. IT MIGHT NOT INTERACT
CC       DIRECTLY WITH DNA BUT WITH DNA-BOUND PROTEINS.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- SIMILARITY: CONTAINS 10 TPR REPEATS.
CC   -!- SIMILARITY: TO YEAST GAL1 AND CCR4.
CC       -----------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       -----------------------------------------------------------------
DR   EMBL; M23440; AAA34545.1; -.
DR   EMBL; M17826; AAA35103.1; -.
DR   EMBL; X66247; CAA46973.1; -.
DR   EMBL; X78993; CAA55615.1; -.
DR   EMBL; Z35981; CAA85069.1; -.
DR   PIR; S25365; S25365.
```

```
DR   SGD; S000316; CYC8.
DR   InterPro; IPR001440; -.
DR   Pfam; PF00515; TPR; 7.
KW   Transcription regulation; Repressor; Repeat; TPR repeat;
KW   Nuclear protein.
FT   DOMAIN       15     30    POLY-GLN.
FT   REPEAT       46     79    TPR 1.
FT   REPEAT       80    113    TPR 2.
FT   REPEAT      114    147    TPR 3.
FT   REPEAT      150    183    TPR 4.
FT   REPEAT      187    220    TPR 5.
FT   REPEAT      224    257    TPR 6.
FT   REPEAT      258    291    TPR 7.
FT   REPEAT      296    329    TPR 8.
FT   REPEAT      330    363    TPR 9.
FT   REPEAT      364    398    TPR 10.
FT   DOMAIN      493    556    30 X 2 AA TANDEM REPEATS OF Q-A.
FT   DOMAIN      557    587    POLY-GLN.
FT   CONFLICT    547    547    K -> Q (IN REF. 3).
SQ   SEQUENCE   966 AA;   107202 MW;   84B509CF3208C5C0 CRC64;


Query Match             10.6%;
Best Local Similarity   26.2%;   Pred. No. 6.7;
Matches    22;   Conservative   15;   Mismatches   34;   Indels   13;   Gaps   2;
                                      Score 75.5;  DB 1;  Length 966;

Qy    64 QTSITLYSMLLQLNFEM---AAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGF 119
         | .: :|| |: :|.| ||  :|.||:||:|| :|:||:|  .:|:|||| .
Db    97 QRAAELYERALLVNPELSDVWATLGHCYLMLDDLQRAYNAYQQALYHLSNPNVPKLWHGI 156

Qy   120 -------IEHAAEVREYIAQLD 134
                 ::| | :||
Db   157 GILYDRYGSLDYAEEAFAKVLELD 180


RESULT  3
DPA4_BPR69
ID   DPA4_BPR69     STANDARD;     PRT;    320 AA.
AC   O64301;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA POLYMERASE ACCESSORY PROTEIN 44 (PROTEIN GP44) (CLAMP LOADER LARGE
DE   SUBUNIT).
GN   44.
OS   Bacteriophage RB69.
OC   Viruses; dsDNA viruses, no RNA stage; Tailed phages; Myoviridae;
OC   T4-like phages.
OX   NCBI_TaxID=12353;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98215164; PubMed=9555879;
RA   Yeh L.-S., Hsu T., Karam J.D.;
RT   "Divergence of a DNA replication gene cluster in the T4-related
RT   bacteriophage RB69.";
RL   J. Bacteriol. 180:2005-2013(1998).
CC   -!- FUNCTION: SLIDING CLAMP (GP45) LOADER. REQUIRED FOR ELONGATION OF
CC       PRIMED TEMPLATES BY DNA POLYMERASE. POSSESSES DNA-DEPENDENT ATPASE
CC       ACTIVITY ON ITS OWN AND WITHIN THE HETERODIMER GP44/GP62.
CC   -!- SUBUNIT: HETERODIMER OF GP44 AND GP62 THAT FORMS A COMPLEX WITH
CC       GP45 IN THE PRESENCE OF ATP.
CC   -!- SIMILARITY: PARTIAL TO E.COLI DNA POLYMERASE III SUBUNITS
CC       GAMMA/TAU AND TO HUMAN ACTIVATOR 1, 37 AND 40 KDA SUBUNITS.
CC       -----------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC       -----------------------------------------------------------------
DR   EMBL; AF039565; AAC39311.1; -.
```

```
DR   InterPro; IPR001939; -.
DR   Pfam; PF00004; AAA; 1.
KW   DNA replication; ATP-binding.
FT   NP_BIND       50     57       ATP (POTENTIAL).
SQ   SEQUENCE   320 AA;  36042 MW;  3EB83BE9C4BDB609 CRC64;

Query Match             10.3%;  Score 73.5;  DB 1;  Length 320;
Best Local Similarity   20.7%;  Pred. No. 2.9;
Matches   29;  Conservative   23;  Mismatches   49;  Indels   39;  Gaps    4;

Qy      1 MTSSQQRVERFLQYFSAGCKTPIHLK-DGVCALYNEQDEEAAVLEVPQHSDSLLLHCRI 58
Db    114 LAESQRHLRTFMEEFSSNCSIITANNIDGII------------EPLRSRCRV 154

Qy     59 IEADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSG 118
Db    155 IEFGRPPEEDKISMMKKMIHRM------------VEICKNEIEIADMKVVAALVK - 198

Qy    119 FIEHAAEVREYIAQLDESSA 138
Db    199 --KNFPDFRRTIGQLDQYSS 216


RESULT   4
EPA1_HUMAN
ID   EPA1_HUMAN     STANDARD;      PRT;     976 AA.
AC   P21709; Q15405;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   EPHRIN TYPE-A RECEPTOR 1 PRECURSOR (EC 2.7.1.112) (TYROSINE-PROTEIN
DE   KINASE RECEPTOR EPH).
GN   EPHA1 OR EPHT1 OR EPHT OR EPH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88070650; PubMed=2825356;
RA   Hirai H., Maru Y., Hagiwara K., Nishida J., Takaku F.;
RT   "A novel putative tyrosine kinase receptor encoded by the eph gene.";
RL   Science 238:1717-1720(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9299440; PubMed=10369740;
RA   Owshalimpur D., Kelley M.J.;
RT   "Genomic structure of the EPHA1 receptor tyrosine kinase gene.";
RL   Mol. Cell. Probes 13:169-173(1999).
RN   [3]
RP   SEQUENCE OF 286-976 FROM N.A.
RC   TISSUE=Placenta;
RA   Tuzi N.L.;
RL   Submitted (MAY-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: RECEPTOR FOR MEMBERS OF THE EPHRIN-A FAMILY. BINDS WITH
CC       A LOW AFFINITY TO EPHRIN-A1.
CC   -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC       PROTEIN TYROSINE PHOSPHATE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: OVEREXPRESSED IN SEVERAL CARCINOMAS.
CC   -!- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. EPHRIN
CC       RECEPTOR SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).


DR   EMBL; M18391; AAA36747.1; ALT_SEQ.
DR   EMBL; AF101171; AAD43440.1; JOINED.
DR   EMBL; AF101165; AAD43440.1; JOINED.
DR   EMBL; AF101166; AAD43440.1; JOINED.
DR   EMBL; AF101167; AAD43440.1; JOINED.
DR   EMBL; AF101168; AAD43440.1; JOINED.
DR   EMBL; AF101169; AAD43440.1; JOINED.
DR   EMBL; AF101170; AAD43440.1; JOINED.
DR   EMBL; 227409; CAA81796.1; -.
DR   PIR; A34076; A34076.
DR   HSSP; P00523; 2PTK.
DR   MIM; 179610; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR000719; -.
DR   InterPro; IPR001090; -.
DR   InterPro; IPR001245; -.
DR   InterPro; IPR001426; -.
DR   InterPro; IPR001660; -.
DR   InterPro; IPR001777; -.
DR   Pfam; PF01404; EPH_lbd; 1.
DR   Pfam; PF00536; SAM; 1.
DR   Pfam; PF00041; fn3; 2.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00790; RECEPTOR_TYR_KIN_V_1; 1.
DR   PROSITE; PS00791; RECEPTOR_TYR_KIN_V_2; 1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW   Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW   Receptor; Transmembrane; Glycoprotein; Signal; Repeat.
FT   SIGNAL        1     23
FT   CHAIN        24    976       POTENTIAL.
FT   DOMAIN       24    547       EPHRIN TYPE-A RECEPTOR 1.
FT   TRANSMEM    548    568       EXTRACELLULAR (POTENTIAL).
FT   DOMAIN      569    976       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      191    329       CYS-RICH.
FT   DOMAIN      330    442       FIBRONECTIN TYPE-III.
FT   DOMAIN      443    537       FIBRONECTIN TYPE-III.
FT   DOMAIN      624    884       PROTEIN KINASE.
FT   DOMAIN      911    976       SAM.
FT   SITE        974    976       PDZ-BINDING MOTIF (POTENTIAL).
FT   NP_BIND     630    638       ATP (BY SIMILARITY).
FT   BINDING     656    656       ATP (BY SIMILARITY).
FT   ACT_SITE    749    749       BY SIMILARITY.
FT   MOD_RES     599    599       PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT   MOD_RES     605    605       PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT   MOD_RES     781    781       PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT   MOD_RES     930    930       PHOSPHORYLATION (AUTO-) (POTENTIAL).
FT   CARBOHYD    338    338       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    414    414       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT    398    398       A -> G (IN REF. 1).
FT   CONFLICT    581    616       QRDRATDVDREDKlWLKPYVDLQAYEDPAQGALDFT ->
FT                                HVTAPPMWIERTSCAEALCGTSRHTTRTLHREPWTLPGGWSN
FT                                FPS (IN REF. 1).
SQ   SEQUENCE   976 AA;  108080 MW;  50512CCF13F03808 CRC64;

Query Match             10.3%;  Score 73.5;  DB 1;  Length 976;
Best Local Similarity   27.5%;  Pred. No. 11;
Matches   36;  Conservative   20;  Mismatches   52;  Indels   23;  Gaps    6;

Qy      7 RVERFLQYFSAGCKT------PIHLKDGVCAlYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
Db     94 RVHVELQFTVRDCKSFPGGAGPLGCKETFNLLYMESDQDVGI----QLRRPLFQKVTTVA 149

Qy     61 ADPQTSI-TLYSMLLQLNFEMAAM------RGCWLALDE------LHNVRLCFQQSLEHLD 108
Db    150 ADQSFTIRDLASGSVKLNVERCSLGRLTRRGLYLAFHNPGACVALVSVRVFYQRCPETLN 209
```

```
Qy   24 HLKDGVCALYNEQDEEAAVLEV----------PQHSDSL--LLHCRIIEADPQTSITLY 70
           |||  ||                            | |||  ||||  ||
Db  165 HFPDAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLY 224

Qy   71 SMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEA----SFSDIVSGFIEHAAE 125
            ||  ||  ||  |  |  |:| ||  :|:|       ||    :|:|
Db  225 RKALEVFPEFAAAHS---NLASVLQQQGKLQEALMHYKEAIRISPFFADAYSNMGNTLKE 281

Qy  126 VRE 128
            :::
Db  282 MQD 284
```

RESULT 6
OGT1_RAT
ID   OGT1_RAT   STANDARD;   PRT;   1036 AA.
AC   P56558;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--PEPTIDE N-ACETYLGLUCOSAMINYLTRANSFERASE 110
DE   KDA SUBUNIT (EC 2.4.1.-) (O-GLCNAC TRANSFERASE P110 SUBUNIT).
GN   OGT.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
RX   MEDLINE=97238869; PubMed=9083067;
RA   Kreppel L.K., Blomberg M.A., Hart G.W.;
RT   "Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and
RT   characterization of a unique O-GlcNAc transferase with multiple
RT   tetratricopeptide repeats.";
RL   J. Biol. Chem. 272:9308-9315(1997).
CC   -!- FUNCTION: ADDITION OF NUCLEOTIDE-ACTIVATED SUGARS DIRECTLY ONTO
CC   THE POLYPEPTIDE THROUGH O-GLYCOSIDIC LINKAGE WITH THE HYDROXYL OF
CC   SERINE OR THREONINE.
CC   -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + PEPTIDE = UDP +
CC   N-ACETYL-BETA-D-GLUCOSAMINYL-PEPTIDE.
CC   -!- ENZYME REGULATION: BY TYROSINE PHOSPHORYLATION AND O-GLCNAC
CC   MODIFICATIONS.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBUNIT: HETEROTRIMER OF TWO 110 KDA AND ONE 78 KDA SUBUNITS. IT
CC   IS NOT KNOWN IF THE 78 KDA SUBUNIT IS ENCODED BY A SEPARATE GENE
CC   OR IS THE PRODUCT OF EITHER OF A PROTEOLYTIC DEGRADATION OR AN
CC   ALTERNATIVE INITIATION OF THE 110 KDA SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC (POSSIBLE).
CC   -!- TISSUE SPECIFICITY: APPEARS TO BE PRESENT IN ALL TISSUES EXAMINED
CC   EXCEPT KIDNEY.
CC   -!- PTM: AUTOMODIFIED BY O-GLYCOSYLATION WITH O-GLCNAC.
CC   -!- SIMILARITY: CONTAINS 12 TPR REPEATS.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; U76557; AAC53121.1; -.
DR   InterPro; IPR001440; -.
DR   Pfam; PF00515; TPR; 11.
KW   Transferase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat;
KW   Phosphorylation; Glycoprotein.
FT   REPEAT   79   112   TPR 1.
FT   REPEAT   113   146   TPR 2.
FT   REPEAT   147   180   TPR 3.
FT   REPEAT   181   214   TPR 4.
FT   REPEAT   215   248   TPR 5.

```
Qy  109 E-ASFSDIVSG 118
           | ||  |:|
Db  210 GLAQFPDTLPG 220
```

RESULT 5
OGT1_HUMAN
ID   OGT1_HUMAN   STANDARD;   PRT;   920 AA.
AC   O15294;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE--PEPTIDE N-ACETYLGLUCOSAMINYLTRANSFERASE 100
DE   KDA SUBUNIT (EC 2.4.1.-) (O-GLCNAC TRANSFERASE P100 SUBUNIT).
GN   OGT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 91-110 AND 829-845.
RC   TISSUE=Liver;
RX   MEDLINE=97238870; PubMed=9083068;
RA   Lubas W.A., Frank D.W., Krause M., Hanover J.A.;
RT   "O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein
RT   containing tetratricopeptide repeats.";
RL   J. Biol. Chem. 272:9316-9324(1997).
CC   -!- FUNCTION: ADDITION OF NUCLEOTIDE-ACTIVATED SUGARS DIRECTLY ONTO
CC   THE POLYPEPTIDE THROUGH O-GLYCOSIDIC LINKAGE WITH THE HYDROXYL OF
CC   SERINE OR THREONINE.
CC   -!- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + PEPTIDE = UDP +
CC   N-ACETYL-BETA-D-GLUCOSAMINYL-PEPTIDE.
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBUNIT: HETEROTRIMER OF TWO 100 KDA AND ONE 70 KDA SUBUNITS. IT
CC   IS NOT KNOWN IF THE 70 KDA SUBUNIT IS ENCODED BY A SEPARATE GENE
CC   OR IS THE PRODUCT OF EITHER OF A PROTEOLYTIC DEGRADATION OR AN
CC   ALTERNATIVE INITIATION OF THE 100 KDA SUBUNIT (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR AND CYTOPLASMIC (POSSIBLE).
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PANCREAS AND TO A LESSER
CC   EXTENT IN SKELETAL MUSCLE, HEART, BRAIN AND PLACENTA. PRESENT IN
CC   TRACE AMOUNTS IN LUNG AND LIVER.
CC   -!- SIMILARITY: CONTAINS 10 TPR REPEATS.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; U77413; AAB63466.1; -.
DR   MIM; 300255; -.
DR   InterPro; IPR001440; -.
DR   Pfam; PF00515; TPR; 8.
KW   Transferase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat.
FT   REPEAT   35   64   TPR 1.
FT   REPEAT   65   98   TPR 2.
FT   REPEAT   99   132   TPR 3.
FT   REPEAT   133   166   TPR 4.
FT   REPEAT   167   200   TPR 5.
FT   REPEAT   201   234   TPR 6.
FT   REPEAT   245   268   TPR 7.
FT   REPEAT   269   302   TPR 8.
FT   REPEAT   303   336   TPR 9.
FT   REPEAT   337   347   TPR 10 (INCOMPLETE).
FT   DOMAIN   361   377   NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ   SEQUENCE   920 AA; 103011 MW; 766BF416ABD547C4 CRC64;

Query Match   10.0%; Score 71.5; DB 1; Length 920;
Best Local Similarity   24.4%; Pred. No. 16;
Matches   30; Conservative   21; Mismatches   51; Indels   21; Gaps 4;
```

```
FT   REPEAT       249    282       TPR 6.
FT   REPEAT       283    316       TPR 7.
FT   REPEAT       317    350       TPR 8.
FT   REPEAT       351    384       TPR 9.
FT   REPEAT       385    418       TPR 10.
FT   REPEAT       419    452       TPR 11.
FT   REPEAT       453    463       TPR 12 (INCOMPLETE).
FT   DOMAIN       478    493       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   MOD_RES      979    979       PHOSPHORYLATION (PROBABLE).
SQ   SEQUENCE   1036 AA;  115605 MW;  3F057CABDD019BD6 CRC64;

Query Match
Best Local Similarity   10.0%;    Score 71.5;  DB 1;  Length 1036;
Matches   30;  Conservative   21;  Mismatches  21;  Indels  51;  Gaps  4;

Qy   24  HLKDGVCALYNEQDEEAAVLEV----------PQHSDSL-LLHCRIIEADPQTSITLY  70
             |:|||            |:|||  : ::;:;; ||
Db  281  HFPDAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQNIEEAVRLY 340

Qy   71  SMLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEA---SFSDIVSGFIEHAAE  125
             |:|| | ||          |   | : :|| |  |
Db  341  RKALEVFPEFAAAHS---NLASVLQQQGKLQEALMHYKEATRISPTFADAYSNMGNTLKE 397

Qy  126  VRE 128
             :::
Db  398  MQD 400

RESULT   7
NEL2_MOUSE
ID   NEL2_MOUSE     STANDARD;      PRT;    816 AA.
AC   Q61220;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   PROTEIN KINASE C-BINDING PROTEIN NELL2 PRECURSOR (NEL-LIKE PROTEIN 2)
DE   (MEL91 PROTEIN).
GN   NELL2 OR MEL91.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Elkins D.A., Rossi J.;
RL   Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   TSP N-TERMINAL DOMAIN.
RX   MEDLINE=98153258; PubMed=9480764;
RA   Beckmann G., Hanke J., Bork P., Reich J.;
RT   "Merging extracellular domains: fold prediction for laminin G-like
RT   and amino-terminal thrombospondin-like modules based on homology to
RT   pentraxins.";
RL   J. Mol. Biol. 275:725-730(1998).
CC   -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 VWFC DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; U59230; AAB02924.1; ALT_INIT.
DR   MGD; MGI:1858510; Nell2.
DR   HSSP; P00740; 1IXA.
DR   InterPro; IPR000152; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001007; -.
DR   InterPro; IPR001881; -.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF00093; vwc; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01208; VWFC; 2.
KW   Glycoprotein; EGF-like domain; Repeat; Signal.
FT   SIGNAL         1     24       POTENTIAL.
FT   CHAIN         25    816       PROTEIN KINASE C-BINDING PROTEIN NELL2.
FT   DOMAIN        30    258       TSP N-TERMINAL.
FT   DOMAIN       272    331       VWFC 1.
FT   DOMAIN       332    396       VWFC 2.
FT   DOMAIN       397    439       EGF-LIKE 1.
FT   DOMAIN       440    481       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       482    522       EGF-LIKE 3.
FT   DOMAIN       521    553       EGF-LIKE 4.
FT   DOMAIN       555    601       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       602    637       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN       638    693       VWFC 3.
FT   DOMAIN       698    756       VWFC 4.
FT   DOMAIN       758    813       VWFC 5.
FT   DISULFID     401    413       BY SIMILARITY.
FT   DISULFID     407    422       BY SIMILARITY.
FT   DISULFID     424    438       BY SIMILARITY.
FT   DISULFID     444    457       BY SIMILARITY.
FT   DISULFID     451    466       BY SIMILARITY.
FT   DISULFID     468    480       BY SIMILARITY.
FT   DISULFID     486    499       BY SIMILARITY.
FT   DISULFID     493    508       BY SIMILARITY.
FT   DISULFID     510    521       BY SIMILARITY.
FT   DISULFID     525    535       BY SIMILARITY.
FT   DISULFID     529    541       BY SIMILARITY.
FT   DISULFID     543    552       BY SIMILARITY.
FT   DISULFID     559    572       BY SIMILARITY.
FT   DISULFID     566    581       BY SIMILARITY.
FT   DISULFID     583    600       BY SIMILARITY.
FT   DISULFID     606    619       BY SIMILARITY.
FT   DISULFID     613    628       BY SIMILARITY.
FT   DISULFID     630    636       BY SIMILARITY.
FT   CARBOHYD      53     53       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     225    225       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     293    293       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     298    298       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     517    517       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     615    615       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     635    635       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE    816 AA;  91163 MW;  5BD00A946F87E74D CRC64;

Query Match                        9.9%;    Score 71;  DB 1;  Length 816;
Best Local Similarity   24.68%;   Pred. No. 16;
Matches   33;  Conservative   12;  Mismatches  55;  Indels  34;  Gaps  4;

Qy    3  SSQQRVERFLQ----------YFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVQHSDSLLLH  55
            :  ||       |||||
Db   69  ASTATAERFLQKLRNKHEFTILVTLKQIHLNSGVILSIHHLDHRYLELESSGHRNEIRLH 128

Qy   56  CRIIEADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQS--LEHLD---- 108
            |  | : :   ||          | ||||:|   | : :|| :|:|
Db  129  YRSGTHRPHTEVPYI---------------------LADAKWHKLSLAFSASHLIHIDCNKIY 172

Qy  109  ---EASFSDIVSG 118
Db  173  ERVVEMPFTDLALG 186

RESULT   8
POLG_EC09B
ID   POLG_EC09B     STANDARD;      PRT;   2203 AA.
AC   Q66577;
```

```
Query Match               9.9%;  Score 70.5;  DB 1;  Length 2203;
Best Local Similarity    24.3%;  Pred. No. 55;
Matches   33;  Conservative  17;  Mismatches  57;  Indels  29;  Gaps   5;

Qy   12 LQYFSAGCKTPIHLKD---GVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSIT  68
          : ||   : :| |     |  ::  .    :       :|
Db  463 LAYSHAGASVPKSRRDAMLGTHVIWDVGLQSSCVLCVPWISQT--HYRLVAQDEYTS-- 517

Qy   69 LYSMLLQLNFEMAAMRGCWLALDEL------HNVRLCFQQSLEHLDEASFSDIVSGFIE 121
          |:|    :        |  : |        ||| |    |  |:|
Db  518 -----AGYITCWQTNIVVPPETPSDCVVLCFVSACNDFSVRMLKD--TPFIE 563

Qy  122 HAAEVREYIAQLDESS 137
          |||  :    : : :
Db  564 QAAELQNDVRQAVEGA 579
```

RESULT 9
AMPN_PLUXY
ID   AMPN_PLUXY   STANDARD;   PRT;   946 AA.
AC   P91087;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   AMINOPEPTIDASE N PRECURSOR (EC 3.4.11.2) (MICROSOMAL AMINOPEPTIDASE)
DE   (APN1).
GN   APN1.
OS   Plutella xylostella (Diamondback moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Yponomeutoidea; Plutellidae; Plutella.
OX   NCBI_TaxID=51655;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   TISSUE=Midgut;
RX   MEDLINE=98000273; PubMed=9342226;
RA   Denolf P.H., Hendrickx K., van Damme J., Jansens S., Peferoen M.,
RA   Degheele D., van Rie J.;
RT   "Cloning and characterization of Manduca sexta and Plutella
RT   xylostella midgut aminopeptidase N related to Bacillus thuringiensis
RT   toxin-binding proteins.";
RL   Eur. J. Biochem. 248:748-761(1997).
CC   -!- CATALYTIC ACTIVITY: AMINOACYL-PEPTIDE + H(2)O = AMINO ACID +
CC   OLIGOPEPTIDE.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLOPROTEASE);
CC   ALSO KNOWN AS THE PEPN SUBFAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X97878; CAA66467.1; -.
DR   MEROPS; M01.013; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001930; -.
DR   Pfam; PF01433; Peptidase_M1; 1.
DR   PRINTS; PR00756; ALADIPTASE.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Aminopeptidase; Zinc; Glycoprotein;
KW   GPI-anchor; Signal.
FT   SIGNAL   1   15   POTENTIAL.
FT   CHAIN   16   ?   AMINOPEPTIDASE N.
FT   PROPEP   ?   946   REMOVED IN MATURE FORM (POTENTIAL).
FT   METAL   344   344   ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   345   345   BY SIMILARITY.
FT   METAL   348   348   ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL   367   367   ZINC (CATALYTIC) (BY SIMILARITY).

DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   GENOME POLYPROTEIN [CONTAINS: COAT PROTEIN VP4 (P1A); COAT PROTEIN VP2
DE   (P1B); COAT PROTEIN VP3 (P1C); COAT PROTEIN VP1 (P1D); PICORNAIN 2A
DE   (EC 3.4.22.29) (P2A); CORE PROTEIN P2B; CORE PROTEIN P2C; CORE PROTEIN
DE   P3A; GENOME-LINKED PROTEIN VPG (P3B); PICORNAIN 3C (EC 3.4.22.28)
DE   (PROTEASE 3C) (P3C); RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (P3D)].
OS   Echovirus 9 (strain Barty).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Enterovirus.
OX   NCBI_TaxID=103914;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97033384; PubMed=8879131;
RA   Zimmermann H., Eggers H.J., Nelsen-Salz B.;
RT   "Molecular cloning and sequence determination of the complete genome
RT   of the virulent echovirus 9 strain Barty.";
RL   Virus Genes 12:149-154(1996).
CC   -!- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE
CC   AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE
CC   PROTEASES.
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC   EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC   VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS
CC   CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.
CC   -!- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   -!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X92886; CAA63480.1; -.
DR   InterPro; IPR000081; -.
DR   InterPro; IPR000199; -.
DR   InterPro; IPR000605; -.
DR   InterPro; IPR001205; -.
DR   InterPro; IPR001643; -.
DR   InterPro; IPR001676; -.
DR   InterPro; IPR002527; -.
DR   Pfam; PF00548; Cys-protease-3C; 1.
DR   Pfam; PF00947; Pico_P2A; 1.
DR   Pfam; PF01552; Pico_P2B; 1.
DR   Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR   Pfam; PF00910; RNA_helicase; 1.
DR   Pfam; PF00073; rhv; 3.
DR   PRINTS; PR00918; CALICVIRUSNS.
KW   Polyprotein; Coat protein; Core protein; Transferase; Myristate;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease.
FT   CHAIN   2   69   COAT PROTEIN VP4.
FT   CHAIN   70   330   COAT PROTEIN VP2.
FT   CHAIN   331   569   COAT PROTEIN VP3.
FT   CHAIN   570   869   COAT PROTEIN VP1.
FT   CHAIN   870   1019   PICORNAIN 2A.
FT   CHAIN   1020   1118   CORE PROTEIN P2B.
FT   CHAIN   1119   1447   CORE PROTEIN P2C.
FT   CHAIN   1448   1536   CORE PROTEIN P3A.
FT   CHAIN   1537   1558   GENOME-LINKED PROTEIN VPG.
FT   CHAIN   1559   1741   PICORNAIN 3C.
FT   CHAIN   1742   2203   RNA-DIRECTED RNA POLYMERASE.
FT   LIPID   2   2   MYRISTATE (BY SIMILARITY).
FT   ACT_SITE   1705   1705   PROTEASE 3C (POTENTIAL).
FT   ACT_SITE   1719   1719   PROTEASE 3C (POTENTIAL).
SQ   SEQUENCE   2203 AA; 246279 MW; 79B78F8C99D02777 CRC64;
```

```
FT   ACT_SITE    429    429        PROTON DONOR (POTENTIAL).
FT   CARBOHYD     60     60        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    550    550        N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    605    605        N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE    946 AA;  106577 MW;  1DFDB1A364067BFB CRC64;

Query Match             9.8%;  Score 70; DB 1; Length 946;
Best Local Similarity  20.6%;  Pred. No. 23;
  Matches   33;  Conservative   25;  Mismatches   62;  Indels   40;  Gaps    5;

Qy   19 CKTPIHLKDGVCALYNEQDE----------EAAVLEVPQHSDSL--LLHCRIIEADPQ 64
           | |:| :||||| |:|                |::| |:  |    :||| |:: |
Db   14 CGNPVQLTDNSIALQNTYDNVVLPGESFPTFYDVQLFFDPEYEASFNGTVAIRVVPRIAT 73

Qy   65 TSITLYSMLLQL----------NFEMAAMRGCWLALDELHNVRLCFQQSLEHLD--- 108
           | |:|:||||:                |||||||||||| :| :  | |:||  | |
Db   74 QEIVLHAMEMEILSIRAYSDLPSDDNLNENLFSSYTLATDDTHLLKIQFTRVLDALQPIT 133

Qy  109 -EASFSD---------IVSGFIEHAAEVREYIAQLDESSA 138
            :|:||          : :|| ||: |: |||:
Db  134 VEISYSAQYAPNMFGVYVSRYVENGATVSLVTSQLQPTFA 173


RESULT 10
RRPL_MEASA
ID   RRPL_MEASA    STANDARD;    PRT;  2183 AA.
AC   P35975;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE   (L PROTEIN).
GN   L.
OS   Measles virus (strain AIK-C) (Subacute sclerose panencephalitis
OS   virus).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Morbillivirus.
OX   NCBI_TaxID=36408;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93227570; PubMed=8470368;
RA   Mori T., Sasaki K., Hashimoto H., Makino S.;
RT   "Molecular cloning and complete nucleotide sequence of genomic RNA of
RT   the AIK-C strain of attenuated measles virus.";
RL   Virus Genes 7:67-81(1993).
CC   -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC       FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC       SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS, RNA EDITING OF THE P
CC       GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC   -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; S58435; AAB26147.1; -.
DR   PIR; G48556; G48556.
DR   InterPro; IPR001016; -.
DR   Pfam; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2183 AA;  247754 MW;  12A6686F4CA62199 CRC64;

Query Match             9.7%;  Score 69.5; DB 1; Length 2183;
Best Local Similarity  20.0%;  Pred. No. 69;
  Matches   38;  Conservative   26;  Mismatches   63;  Indels   63;  Gaps    7;

Qy    4 SQQRVERFLQYFSAGCK------------TPIHLKDGVCALYNEQDEEAAVLEVP 46
```

```
Db  158 SSQWFEPFLFWFTVKTEMRSVIKSQTHTCHRRRHTPVFFTGSSVELLISRDLVAIISKES 217
                | | | || :|: ||:
Qy   47 QH-----SDSLLHCRIEAD--PQTSITLYSMLLQLNFEMAAMRGCWLALD------- 91
               :       ::| :|::      | | ||| :||| |||||||||||:|
Db  218 QHVYYLTFELVLMYCDVIGGRLMTETAMTIDARYTEL--LGRVRYMWKLIDGFFPALGN 274

Qy   92 ----------------------ELHNVRL-CFQQSLEHLDEASFSDIVSGFIEHAAE 125
                                  ||  :|  |||||||:||||   ||||:| ||||
Db  275 PTYQIVAMLEPLSLAYLQLRDITVELRGAFLNHCFTEIHDVLDQNGFSD--EGTYHELIE 332

Qy  126 VREYIAQLDE 135
             |:||:
Db  333 ALDYIFTTDD 342


RESULT 11
RRPL_MEASE
ID   RRPL_MEASE    STANDARD;    PRT;  2183 AA.
AC   P12576;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE   (L PROTEIN).
GN   L.
OS   Measles virus (strain Edmonston) (Subacute sclerose panencephalitis
OS   virus).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC   Paramyxoviridae; Paramyxovirinae; Morbillivirus.
OX   NCBI_TaxID=11235;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88219537; PubMed=2835864;
RA   Blumberg B.M., Crowley J.C., Silverman J.I., Menonna J., Cook S.D.,
RA   Dowling P.C.;
RT   "Measles virus L protein evidences elements of ancestral RNA
RT   polymerase.";
RL   Virology 164:487-497(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90085790; PubMed=2596022;
RA   Cattaneo R., Schmid A., Spielhofer P., Kaelin K., Baczko K.,
RA   Meulen V., Pardowitz J., Flanagan S., Rima B.K., Udem S.A.;
RT   "Mutated and hypermutated genes of persistent measles viruses which
RT   caused lethal human brain diseases.";
RL   Virology 173:415-425(1989).
CC   -!- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC       FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC       SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAS. RNA EDITING OF THE P
CC       GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC   -!- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; M20865; AAA46430.1; -.
DR   EMBL; K01711; AAA75501.1; -.
DR   PIR; A28919; ZLNZMV.
DR   InterPro; IPR001016; -.
DR   Pfam; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2183 AA;  247646 MW;  1B0B03CA2E2B6EA5 CRC64;

Query Match             9.7%;  Score 69.5; DB 1; Length 2183;
Best Local Similarity  20.0%;  Pred. No. 69;
  Matches   38;  Conservative   26;  Mismatches   63;  Indels   63;  Gaps    7;
```

```
Qy     4  SQQRVERFLQYFSAGCK----------------------TPIHLKDGVCALYNEQDEEAAVLEVP  46
          |  |-|| |.|: .|                         ||.|: |||| |||||| |   ||
Db   231  KLP-HFEEGAVTVQDLSAQWAATLLEPKNEEWILDACAAPGGKTHILELAPQANV----  285

Qy   158  SSQWFEPFLFWFTVKTEMRSVIKSQTHTCHRRRHTPVFTGSSVELLISRDLVAIISKES  217
          :|||.|::  ||| :::|: :||:.|||| :||:| .||:  :::||||:  :||| |
Db    72  MLLQLNFEMAAMRGCWLALD-ELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYI  130

Qy    47  QH-----SDSLLHCRIIEAD-PQTSITLYSMLLQLNFEMAAMRGCWLALD----------  91
          ||      :|||:| :|::| .|:|:|   |.:.|||||||||||||||
Db   286            -IALDVESHRLKR-VEENLERLNQQAI--VVCG--DASKPDEWL  323

Qy   218  QHVYYLTFELVLMYCDVIEGRLMTETAMTIDARYTEL---LGRVRYMWKLIDGFFPALGN  274
          || |                   ||  | | |:::|       : | ||||:||||
Db   131  AQLDESSA  138

Qy    92  ----------ELHNVRL-CFQQSLEHLDEASFSDIVSGFIEHAAE  125
                    |||:||| ||||||||||||||||||||||||||
Db   324  AEIGKSAA  331

Qy   275  PTYQIVAMLEPLSLAYLQLRDITVELRGAFLNHCFTEIHDVLDQNGFSD--EGTYHELIE  332
          ||: |  |||  :| |: ||  :   :|||||||||||||||||||     ||||||:|

Qy   126  VREYIAQLDE  135
          :|: ||
Db   333  ALDYIFITDD  342
```

RESULT 12
```
SUN_HAEIN
ID   SUN_HAEIN               STANDARD;       PRT;      451 AA.
AC   P44788;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   SUN PROTEIN (FMU PROTEIN).
GN   SUN OR FMU OR HI0624.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- SIMILARITY: BELONGS TO THE SUN (BACTERIAL) / NUCLEOLAR PROTEIN
CC       NOL1/NOP2 (EUKARYOTES) FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U32745; AAC22284.1; -.
DR   TIGR; HI0624; -.
DR   InterPro; IPR000139; -.
DR   InterPro; IPR001678; -.
DR   Pfam; PF01189; NOl1_Nop2_Sun; 1.
DR   Pfam; PF01029; NusB; 1.
DR   PROSITE; PS01153; NOl1_NOP2_SUN; 1.
SQ   SEQUENCE  451 AA;  50597 MW;  D91FAB88FFDE34B0 CRC64;

Query Match             9.7%;  Score 69;  DB 1;  Length 451;
Best Local Similarity   22.7%;  Pred. No. 12;
Matches  29;  Conservative  30;  Mismatches  33;  Indels  36;  Gaps  7;

Qy    20  KTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHC-------RIIEADPQTSITLYS  71
```

```
RESULT 13
MERL_HUMAN
ID   MERL_HUMAN              STANDARD;       PRT;      595 AA.
AC   P35240;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   MERLIN (SCHWANNOMIN) (NEUROFIBROMIN 2).
GN   NF2 OR SCH.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93201601; PubMed=8453669;
RA   Trofatter J.A., MacCollin M.M., Rutter J.L., Murrell J.R., Duyao M.P.,
RA   Parry D.N., Eldridge R., Kley N., Menon A.G., Pulaski K., Haase V.H.,
RA   Ambrose C.M., Munroe D., Bove C., Haines J.L., Martuza R.L.,
RA   Macdonald M.E., Seizinger B.R., Short M.P., Buckler A.J.,
RA   Gusella J.F.;
RT   "A novel moesin-, ezrin-, radixin-like gene is a candidate for the
RT   neurofibromatosis 2 tumor suppressor.";
RL   Cell 72:791-800(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93281181; PubMed=8379998;
RA   Rouleau G.A., Merel P., Lutchman M., Sanson M., Zucman J.,
RA   Marineau C., Hoang-Xuan K., Demczuk S., Desmaze C., Plougastel B.,
RA   Pulst S., Lenoir G., Bijlsma E., Fashold R., Dumanski J., de Jong P.,
RA   Parry D., Eldrige R., Aurias A., Delattre O., Thomas G.;
RT   "Alteration in a new gene encoding a putative membrane-organizing
RT   protein causes neuro-fibromatosis type 2.";
RL   Nature 363:515-521(1993).
RN   [3]
RP   REVIEW.
RA   Marineau C., Merel P., Rouleau G.A., Thomas G.;
RT   "The gene of neurofibromatosis type 2.";
RL   Medecine/Sciences 11:35-42(1995).
RN   [4]
RP   VARIANT NF2 TYR-220.
RX   MEDLINE=94047514; PubMed=8230593;
RA   MacCollin M., Mohney T., Trofatter J., Wertelecki W., Ramesh V.,
RA   Gusella J.F.;
RT   "DNA diagnosis of neurofibromatosis 2. Altered coding sequence of the
RT   merlin tumor suppressor in an extended pedigree.";
RL   JAMA 270:2316-2320(1993).
RN   [5]
RP   VARIANT NF2 PHE-96 DEL.
RX   MEDLINE=94311312; PubMed=7913580;
RA   MacCollin M., Ramesh V., Jacoby L.B., Louis D.N., Rubio M.-P.,
RA   Pulaski K., Trofatter J.A., Short M.P., Bove C., Eldridge R.,
RA   Parry D.M., Gusella J.F.;
RT   "Mutational analysis of patients with neurofibromatosis 2.";
RL   Am. J. Hum. Genet. 55:314-320(1994).
RN   [6]
RP   VARIANT ARG-46.
RX   MEDLINE=94272478; PubMed=8004107;
RA   Irving R.M., Moffat D.A., Hardy D.G., Barton D.E., Xuereb J.H.,
RA   Maher E.R.;
RT   "Somatic NF2 gene mutations in familial and non-familial vestibular
```

RT   schwannoma.";
RL   Hum. Mol. Genet. 3:347-350(1994).
RN   [7]
RP   VARIANTS MET-219 AND CYS-418.
RX   MEDLINE=94282032; PubMed=8012353;
RA   Jacoby L.B., Maccollin M., Louis D.N., Mohney T., Rubio M.P.,
RA   Pulaski K., Trofatter J.A., Kley N., Seizinger B., Ramesh V.,
RA   Gusella J.F.;
RT   "Exon scanning for mutation of the NF2 gene in schwannomas.";
RL   Hum. Mol. Genet. 3:413-419(1994).
RN   [8]
RP   VARIANTS NF2 SER-62; GLY-106 AND MET-352.
RX   MEDLINE=94362682; PubMed=8081368;
RA   Bourn D., Carter S.A., Mason S., Gareth D., Evans R., Strachan T.;
RT   "Germline mutations in the neurofibromatosis type 2 tumour suppressor
RT   gene.";
RL   Hum. Mol. Genet. 3:813-816(1994).
RN   [9]
RP   VARIANT GLY-79.
RX   MEDLINE=95038740; PubMed=7951231;
RA   Sainz J., Huynh D.P., Figueroa K., Ragge N.K., Baser M.E., Pulst S.M.;
RT   "Mutations of the neurofibromatosis type 2 gene and lack of the gene
RT   product in vestibular schwannomas.";
RL   Hum. Mol. Genet. 3:885-891(1994).
RN   [10]
RP   VARIANTS PHE-273 AND ILE-364.
RX   MEDLINE=94214500; PubMed=8162073;
RA   Bianchi A.B., Hara T., Ramesh V., Gao J., Klein Szanto A.J., Morin F.,
RA   Menon A.G., Trofatter J.A., Gusella J.F., Seizinger B.R., Kley N.;
RT   "Mutations in transcript isoforms of the neurofibromatosis 2 gene in
RT   multiple human tumour types.";
RL   Nat. Genet. 6:185-192(1994).
RN   [11]
RP   VARIANT NF2 PRO-535.
RX   MEDLINE=95395825; PubMed=7666400;
RA   Evans D.G.R., Bourn D., Wallace A., Ramsden R.T., Mitchell J.D.,
RA   Strachan T.;
RT   "Diagnostic issues in a family with late onset type 2
RT   neurofibromatosis.";
RL   J. Med. Genet. 32:470-474(1995).
RN   [12]
RP   VARIANT NF2 PRO-538.
RX   MEDLINE=96163499; PubMed=8566958;
RA   Kluwe L., Mautner V.-F.;
RT   "A missense mutation in the NF2 gene results in moderate and mild
RT   clinical phenotypes of neurofibromatosis type 2.";
RL   Hum. Genet. 97:224-227(1996).
RN   [13]
RP   VARIANTS F-96 DEL; I-117; F-119 DEL; V-122--E-129 DEL AND F-339.
RX   MEDLINE=96209910; PubMed=8655144;
RA   de Vitis L.R., Tedde A., Vitelli F., Ammannati F., Mennonna P.,
RA   Bigozzi U., Montali E., Papi L.;
RT   "Screening for mutations in the neurofibromatosis type 2 (NF2) gene
RT   in sporadic meningiomas.";
RL   Hum. Genet. 97:632-637(1996).
RN   [14]
RP   VARIANT NF2 ARG-234.
RX   MEDLINE=99192373; PubMed=10090912;
RA   Baser M.E., Kluwe L., Mautner V.-F.;
RT   "Germ-line NF2 mutations and disease severity in neurofibromatosis
RT   type 2 patients with retinal abnormalities.";
RL   Am. J. Hum. Genet. 64:1230-1233(1999).
-!- FUNCTION: PROBABLY ACTS AS A MEMBRANE STABILIZING PROTEIN.
-!- TISSUE SPECIFICITY: IN FETAL BRAIN; IN KIDNEY, LUNG, BREAST,
CC   AND OVARY.
-!- DISEASE: NEUROFIBROMATOSIS 2 (NF2) OR CENTRAL NEUROFIBROMATOSIS IS
CC   A GENETIC DISORDER CHARACTERIZED BY BILATERAL VESTIBULAR
CC   SCHWANNOMAS (FORMERLY CALLED ACOUSTIC NEUROMAS), SCHWANNOMAS OF
CC   OTHER CRANIAL AND PERIPHERAL NERVES, MENINGIOMAS, AND EPENDYMOMAS.
CC   IT IS INHERITED IN AN AUTOSOMAL DOMINANT FASHION WITH FULL
CC   PENETRANCE. AFFECTED INDIVIDUALS GENERALLY DEVELOP SYMPTOMS OF
CC   EIGHTH-NERVE DYSFUNCTION IN EARLY ADULTHOOD, INCLUDING DEAFNESS
CC   AND BALANCE DISORDER. ALTHOUGH THE TUMORS OF NF2 ARE

CC   HISTOLOGICALLY BENIGN, THEIR ANATOMIC LOCATION MAKES MANAGEMENT
CC   DIFFICULT, AND PATIENTS SUFFER GREAT MORBIDITY AND MORTALITY.
CC   -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,
CC   RADIXIN, AND TALIN.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L11353; AAA36212.1; -.
DR   EMBL; X72655; CAA51220.1; -.
DR   EMBL; X72656; CAA51220.1; JOINED.
DR   EMBL; X72657; CAA51220.1; JOINED.
DR   EMBL; X72658; CAA51220.1; JOINED.
DR   EMBL; X72659; CAA51220.1; JOINED.
DR   EMBL; X72660; CAA51220.1; JOINED.
DR   EMBL; X72661; CAA51220.1; JOINED.
DR   EMBL; X72662; CAA51220.1; JOINED.
DR   EMBL; X72663; CAA51220.1; JOINED.
DR   EMBL; X72664; CAA51220.1; JOINED.
DR   EMBL; X72665; CAA51220.1; JOINED.
DR   EMBL; X72666; CAA51220.1; JOINED.
DR   EMBL; X72667; CAA51220.1; JOINED.
DR   EMBL; X72668; CAA51220.1; JOINED.
DR   EMBL; X72669; CAA51220.1; JOINED.
DR   EMBL; X72670; CAA51220.1; JOINED.
DR   EMBL; Z22664; CAA80377.1; JOINED.
DR   PIR; S33809; S33809.
DR   MIM; 101000; -.
DR   InterPro; IPR000299; -.
DR   InterPro; IPR000798; -.
DR   Pfam; PF00373; Band_41; 1.
DR   Pfam; PF00769; ERM; 1.
DR   PRINTS; PR00935; BAND41.
DR   PROSITE; PS00660; BAND_41_1; 1.
DR   PROSITE; PS00661; BAND_41_2; 1.
DR   PROSITE; PS50057; BAND_41_3; 1.
KW   Structural protein; Cytoskeleton; Anti-oncogene; Disease mutation.
FT   DOMAIN   74   241   BAND 4.1-LIKE.
FT   VARIANT   46   46   L -> R (IN VESTIBULAR SCHWANNOMA).
FT   /FTId=VAR_000809.
FT   VARIANT   62   62   F -> S (IN NF2).
FT   /FTId=VAR_000810.
FT   VARIANT   79   79   K -> G (IN VESTIBULAR SCHWANNOMA).
FT   /FTId=VAR_000811.
FT   VARIANT   96   96   MISSING (IN NF2, AND SPORADIC
FT   MENINGIOMA).
FT   /FTId=VAR_000812.
FT   VARIANT   106   106   E -> G (IN NF2).
FT   /FTId=VAR_000813.
FT   VARIANT   117   117   L -> I (IN SPORADIC MENINGIOMA).
FT   /FTId=VAR_000814.
FT   VARIANT   119   119   MISSING (IN SPORADIC MENINGIOMA).
FT   /FTId=VAR_000815.
FT   VARIANT   122   129   MISSING (IN SPORADIC MENINGIOMA).
FT   /FTId=VAR_000816.
FT   VARIANT   219   219   V -> M (IN VESTIBULAR SCHWANNOMA).
FT   /FTId=VAR_000817.
FT   VARIANT   220   220   N -> Y (IN NF2).
FT   /FTId=VAR_000818.
FT   VARIANT   234   234   L -> R (IN NF2 AND RETINAL HAMARTOMA;
FT   SEVERE).
FT   /FTId=VAR_009123.
FT   VARIANT   273   273   I -> F (IN BREAST DUCTAL CARCINOMA).

Query Match   9.6%; Score 68.5; DB 1; Length 595;
Best Local Similarity 25.7%; Pred. No. 19;
Matches   27; Conservative 22; Mismatches 43; Indels 13; Gaps   4;

```
Qy    33 YNEQDEEAAVLEVPQHSDSLLLHCRIIE----ADPQTSITLYSMLLQL--NFEMAAMRG 85
         | || ||  | |||  ||| :  :: ||      :  :  :
Db   101 YPENAEEELVQEITQHLFFLQVKKQILDEKIYCPPEASVLLASYAVQAKYGDYDPSVHKR 160

Qy    86 CWLALDELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAEVRE 128
         :|| :|| |    :      ||      | :  ||
Db   161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201

RESULT 14
ID   MERL_MOUSE      STANDARD;       PRT;     596 AA.
AC   P46662;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   MERLIN (SCHWANNOMIN) (NEUROFIBROMIN 2).
GN   NF2 OR NF-2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94282031; PubMed=8012352;
RA   Haase V.H., Trofatter J.A., Maccollin M., Tarttelin E., Gusella J.F.,
RA   Ramesh V.;
RT   "The murine NF2 homologue encodes a highly conserved merlin protein
RT   with alternative forms.";
RL   Hum. Mol. Genet. 3:407-411(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95072570; PubMed=7981675;
RA   Huynh D.P., Nechiporuk T., Pulst S.M.;
RT   "Alternative transcripts in the mouse neurofibromatosis type 2 (NF2)
RT   gene are conserved and code for schwannomins with distinct C-terminal
RT   domains.";
RL   Hum. Mol. Genet. 3:1075-1079(1994).
RN   [3]
RP   SEQUENCE OF 400-596 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=94375073; PubMed=8088840;
RA   Claudio J.O., Marineau C., Rouleau G.A.;
RT   "The mouse neurofibromatosis type 2 gene maps to chromosome 11.";
RL   Genomics 21:437-439(1994).
CC   -!- FUNCTION: PROBABLY ACTS AS A MEMBRANE STABILIZING PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS ARE PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,
CC       RADIXIN, AND TALIN.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; X74671; CAA52737.1; -.
DR   EMBL; L27105; AAA39807.1; -.
DR   EMBL; L27090; AAA63648.1; -.
DR   EMBL; L28176; AAA39808.1; -.
DR   EMBL; X75759; CAA53386.1; -.
DR   MGD; MGI:97307; Nf2.
DR   InterPro; IPR000299; -.
DR   InterPro; IPR000798; -.
DR   Pfam; PF00373; Band_41; 1.
DR   Pfam; PF00769; ERM; 1.
DR   PRINTS; PR00935; BAND41.
DR   PROSITE; PS00660; BAND_41_1; 1.
DR   PROSITE; PS00661; BAND_41_2; 1.
DR   PROSITE; PS50057; BAND_41_3; 1.
KW   Structural protein; Cytoskeleton; Anti-oncogene; Alternative splicing.
FT   DOMAIN       74    241       BAND 4.1-LIKE.
FT   VARSPLIC    581    596       LTLQSAKSRVAFFEEL -> PQAQGRRPICI (IN
FT                                SECOND ISOFORM).
FT   CONFLICT    475    475       I -> T (IN REF. 2).
FT   CONFLICT    554    554       I -> R (IN REF. 2 AND 3).
FT   CONFLICT    570    570       G -> A (IN REF. 2).
SQ   SEQUENCE   596 AA;  69691 MW;  8D06E3B07D2C5851 CRC64;

Query Match             9.6%;  Score 68.5;  DB 1;  Length 596;
Best Local Similarity   25.7%;  Pred. No. 19;
  Matches   27;  Conservative   22;  Mismatches   43;  Indels   13;  Gaps    4;

Qy    33 YNEQDEEAAVLEVPQHSDSLLLHCRIIE----ADPQTSITLYSMLLQL--NFEMAAMRG 85
         | || ||  | |||  ||| :  :: ||      :  :  :
Db   101 YPENAEEELVQEITQHLFFLQVKKQILDEKIYCPPEASVLLASYAVQAKYGDYDPSVHKR 160

Qy    86 CWLALDELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAEVRE 128
         :|| :|| |    :      ||      | :  ||
Db   161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201

RESULT 15
ID   ARDE_CHLMU      STANDARD;       PRT;     478 AA.
AC   P56961;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   SHIKIMATE BIOSYNTHESIS PROTEIN ARODE (INCLUDES: 3-DEHYDROQUINATE
DE   DEHYDRATASE (EC 4.2.1.10) (3-DEHYDROQUINASE) (TYPE I DHQASE);
DE   SHIKIMATE 5-DEHYDROGENASE (EC 1.1.1.25)].
GN   AROE OR AROD/E OR TC0649.
OS   Chlamydia muridarum.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=83560;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MOPN / NIGG;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K.,
RA   Bass S., Linher K., Weidman J., Khouri H., Craven B., Bowman C.,
RA   Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G.,
RA   Salzberg S.L., Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
CC   -!- FUNCTION: BIFUNCTIONAL ENZYME THAT CATALYZES TWO SEQUENTIAL STEPS
CC       OF THE AROMATIC AMINO ACIDS BIOSYNTHETIC PATHWAY. THE FIRST
CC       REACTION IS CATALYZED BY THE 3-DEHYDROQUINATE DEHYDRATASE, CODED
CC       BY THE AROD DOMAIN; THE SECOND REACTION IS CATALYZED BY THE
CC       SHIKIMATE 5-DEHYDROGENASE, CODED BY THE AROE DOMAIN.
CC   -!- CATALYTIC ACTIVITY: 3-DEHYDROQUINATE = 3-DEHYDROSHIKIMATE + H(2)O.
CC   -!- CATALYTIC ACTIVITY: SHIKIMATE + NADP(+) = 5-DEHYDROSHIKIMATE +
CC       NADPH.
CC   -!- PATHWAY: CATALYZES THE THIRD (AROD) AND FOURTH (AROE) STEPS IN THE
CC       BIOSYNTHESIS FROM CHORISMATE OF THE AROMATIC AMINO ACIDS (THE
CC       SHIKIMATE PATHWAY).
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION: BELONGS TO THE TYPE-I 3-
CC       DEHYDROQUINASE FAMILY.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION: BELONGS TO THE SHIKIMATE
CC       DEHYDROGENASE FAMILY.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC       -----------------------------------
DR       EMBL; AE002332; AAF39475.1; -.
DR       TIGR; TC0649; -.
DR       InterPro; IPR001381; -.
DR       InterPro; IPR002907; -.
DR       Pfam; PF01487; DHquinase_I; 1.
DR       Pfam; PF01488; Shikimate_DH; 1.
DR       PROSITE; PS01028; DEHYDROQUINASE_I; FALSE_NEG.
KW       Aromatic amino acid biosynthesis; Lyase; Oxidoreductase; NADP;
KW       Multifunctional enzyme.
FT       DOMAIN        1    208       3-DEHYDROQUINATE DEHYDRATASE.
FT       DOMAIN      209    478       SHIKIMATE 5-DEHYDROGENASE.
FT       ACT_SITE    110              BY SIMILARITY.
FT       ACT_SITE    133              FORMS A SCHIFF-BASE INTERMEDIATE (BY
FT                                    SIMILARITY).
SQ       SEQUENCE   478 AA;  53210 MW;  199EEF5251205AB9 CRC64;


Query Match             9.5%;  Score 68;  DB 1;  Length 478;
Best Local Similarity  26.4%;  Pred. No. 17;
Matches   32;  Conservative   20;  Mismatches   35;  Indels   34;  Gaps    9;

Qy    18 GC------KTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSI---T 68
         ||       :|:||         :  : :||   | | ::::||  ||::::|   |:||
Db   353 GCELLIFNRTKIH-AEAIASRYQAQAFD--IKDLPLHSVSLIINCL-----PPSSIIPQA 404

Qy    69 LYSMLLQLN------FEMAAMRGC------WLALDELHNVRLCF-QQSLEHLDEASF 112
         | |:::::|  : :::|         |   |   |   | | | |   :| | ::|
Db   405 LAPLIVDINTLPKHNSFTQYARLKGCSIIYGHEMFAQQALLQFRLWFPTHSFNHL-EKNF 463

Qy   113 S 113
           |
Db   464 S 464


Search completed: June  5, 2001, 18:23:09
Job time: 365 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:18:11 ; Search time 54.28 Seconds
(without alignments)
175.985 Million cell updates/sec

Title: US-09-596-784-4
Perfect score: 715
Sequence: 1 MTSSQQRVERFLQYFSAGCK.........IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	715	100.0	139	2	T18449	pathogenicity fact
2	273.5	38.3	129	2	T30333	avirulence protein
3	79	11.0	789	2	A70868	probable plsB2 - M
4	78.5	11.0	830	2	C82710	ribonucleoside-dip
5	75.5	10.8	775	2	T45238	probable transfera
6	75.5	10.8	402	2	T19390	hypothetical prote
7	75.5	10.6	966	2	S25365	CYC8 protein - yea
8	73.5	10.3	984	1	A34076	protein-tyrosine k
9	71.5	10.0	1036	2	T31673	N-acetylglucosamin
10	71.5	10.0	2484	2	T26216	hypothetical prote
11	71.5	10.0	2607	2	T26215	hypothetical prote
12	70.5	9.9	1048	2	T31653	hypothetical prote
13	69.5	9.7	291	2	T23051	hypothetical prote
14	69.5	9.7	393	2	S76366	hypothetical prote
15	69.5	9.7	2183	1	2LNZMV	genome polyprotein
16	69.5	9.7	2183	1	G48556	genome polyprotein
17	69	9.7	225	2	F72124	pts iia protein +
18	69	9.7	374	2	A42264	membrane-associate
19	69	9.7	451	1	F64155	hypothetical prote
20	69	9.7	610	2	T02298	probable outer den
21	69	9.7	638	2	T09400	outer dense fiber
22	68.5	9.6	165	2	C72625	hypothetical prote
23	68.5	9.6	167	2	H81717	conserved hypothet
24	68.5	9.6	173	2	S75578	hypothetical prote
25	68.5	9.6	352	2	T44968	gas-vesicle operon
26	68.5	9.6	591	2	I54368	merlin protein - m
27	68.5	9.6	595	2	S33809	neurofibromin 2 -
28	68.5	9.6	596	2	I68664	merlin - mouse
29	68.5	9.6	621	2	T06717	hypothetical prote
30	68.5	9.6	794	2	T46073	hypothetical prote
31	68	9.5	178	2	I39722	ORF12 - Agrobacter
32	68	9.5	478	2	G81679	3-dehydroquinate d
33	68	9.5	638	2	T03791	outer dense fiber
34	68	9.5	923	2	T24712	hypothetical prote
35	68	9.5	1119	2	B70126	surface-located me
36	68	9.5	1144	2	A54810	TMV resistance pro
37	67.5	9.4	167	2	A71563	hypothetical prote
38	67.5	9.4	1123	2	T51517	telomerase reverse
39	67.5	9.4	1141	2	T20611	hypothetical prote
40	67.5	9.4	1196	2	S65245	translation elonga
41	67.5	9.4	1527	2	JE0336	canalicular multis
42	67	9.4	578	2	T15736	hypothetical prote
43	67	9.4	746	2	S66963	hypothetical prote
44	67	9.4	748	2	G83927	alpha-galactosidas
45	67	9.4	921	2	D82513	malT regulatory pr

ALIGNMENTS

RESULT 1
T18449
pathogenicity factor DspB - Erwinia amylovora
C;Species: Erwinia amylovora
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T18449
R;Gaudriault, S.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z18936
A;Accession: T18449
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-139 <GAU>
A;Cross-references: EMBL:Y13831; PIDN:CAA74157.1
A;Experimental source: strain CFBP1430; specific host Pommoideae
C;Genetics:
A;Note: dspB
C;Function:
A;Description: involved in pathogenicity

Query Match   100.0%;   Score 715; DB 2;   Length 139;
Best Local Similarity   100.0%;   Pred. No. 3.8e-64;
Matches 139;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLHCRIIE 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLHCRIIE 60

Qy 61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120

Qy 121 EHAAEVREYIAQLDESSAA 139
        |||||||||||||||||||
Db 121 EHAAEVREYIAQLDESSAA 139

RESULT 2
T30333
avirulence protein - Pseudomonas syringae
C;Species: Pseudomonas syringae
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T30333
R;Bogdanove, A.J.; Kim, J.F.; Wei, Z.; Kolchinsky, P.; Charkowski, A.O.; Conlin, A.K
Proc. Natl. Acad. Sci. U.S.A. 95, 1325-1330, 1998
A;Title: Homology and functional similarity of an hrp-linked pathogenicity locus, dsp
A;Reference number: Z20825; MUID:9815919
A;Accession: T30333
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-129 <BOG>

A;Accession: C82710
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-830 <SIM>
A;Cross-references: GB:AE003954; GB:AE003849; NID:g9106165; PIDN:AAF84006.1; GSPDB:GN
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
,F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa JR., V.E. de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; da Silva Jr, W.A.; da Silv
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1196
C;Superfamily: herpesvirus ribonucleoside-diphosphate reductase large chain

Query Match   11.0%; Score 78.5; DB 2; Length 830;
Best Local Similarity   25.6%; Pred. No. 7.3;
Matches   33; Conservative 21; Mismatches 46; Indels 29; Gaps 6;

Qy   17 AGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRII-EADPQTSITLYSMLLQ 75
         || ||| | | :|| :: :||:| : ||| |||
Db   22 AGSVPPVALS-----VFHPDEPED---EVPIPRQTAMHTRAVEEANVSTWITKEAGNRR 73

Qy   76 LNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEV------- 126
          : |:| :| |||:| |!|:|
Db   74 MPFDCARLE---CAIDTIH------QEFPQLDVAEYKRAVFGFVERKDSVNADDLVDLL 123

Qy   127 -REYIAQLD 134
          || |::|
Db   124 IREAEARVD 132

RESULT 5
T45238
probable transferase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
R;James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z22949
A;Accession: T45238
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-775 <JAM>
A;Cross-references: EMBL:AL049913; PIDN:CAB43153.1
A;Experimental source: cosmid B1610
C;Genetics:
A;Note: plsB

Query Match   10.8%; Score 77.5; DB 2; Length 775;
Best Local Similarity   25.8%; Pred. No. 8.5;
Matches   34; Conservative 16; Mismatches 33; Indels 49; Gaps 5;

Qy   11 FLQYFSAGCKTPIHLKDGVCALYN---------------------------------EQ 36
         || || :|| ::
Db   397 YAAYARGGEKTP----EGVAWLYSFIKAQGERNYGKIYVRPPEAVSMRQYLGAPHGALVQ 452

Qy   37 DEEAAVLEVPQHSDSLLLHCRIIEADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNV 96
          ::|| :|| |||:| || | |: ::| |:|| |
Db   453 DQDAKRLALQKMSFEVA-WRILCATPVTATALVSALL------LTTRGVALTLDQLHHT 504

A;Cross-references: EMBL:O97505; NID:g2978502; PID:g2978504; PIDN:AAC06135.1
C;Genetics:
A;Gene: avrF

Query Match   38.3%; Score 273.5; DB 2; Length 129;
Best Local Similarity   43.1%; Pred. No. 3e-20;
Matches   56; Conservative 27; Mismatches 44; Indels 3; Gaps 3;

Qy   1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
         | ||||||||||:||:| |:|:|:||:|:| | :||||:||||||||||||||||
Db   1 MKTSQPDFARFINSLGAQLGTSLTLQNGVCALYDGQNNEAAIIELPEHSEMVIFPHCRIGR 60

Qy   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
          :| ||||:: || |||||||||||| ||| |:| ::| |||||||||||
Db   61 C-PERAPDLL-RLLSLNFDVARLHGCWFAVDQ-GDVRLCAQRELASLDEPAFCDVTRGFI 117

Qy   121 EHAAEVREYI 130
         ||| ::
Db   118 SQAREARAFL 127

RESULT 3
A70868
probable plsB2 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
R; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: A70868
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-789 <COL>
A;Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16059.1; PID:e123760
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: plsB2

Query Match   11.0%; Score 79; DB 2; Length 789;
Best Local Similarity   24.3%; Pred. No. 6.1;
Matches   34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;

Qy   23 IHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITLYSMLLQLNFEMAA 82
         || || :|| ::|
Db   445 VSMRQYLGAPHGELTQDPAAKRLALQKMSFEVAWRILQATPVTATGLVSALL-----LT 498

Qy   83 MRGCWLALDELHNVRLCFQQSLEHLD-----------------EASFSDIVSGF---- 119
         ||:| |!|:| |: :|
Db   499 TRGTALTLDQLHHT---LQDSLDYLERKQSPVSTSALRLRSREGVRAAADALSNGHPVTR 555

Qy   120 IEHAAEVREYIAQLDESSAA 139
         | ||: || :: ||
Db   556 VDSGREPVWYIAPDDEHAAA 575

RESULT 4
C82710
ribonucleoside-diphosphate reductase alpha chain XF1196 [imported] - Xylella fastidiosa
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
A;Accession: C82710
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below

Qy   97 RLCFQQSLEHLD 108
          |:||:||:|:
Db   505 --LQESLDYILE 513

RESULT 6
T19390
hypothetical protein C18D4.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19390
R;Mortimore, B.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19119
A;Accession: T19390
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-402 <WIL>
A;Cross-references: EMBL:Z81474; PIDN:CAB03900.1; GSPDB:GN00023; CESP:C18D4.2
A;Experimental source: clone C18D4
C;Genetics:
A;Gene: CESP:C18D4.2
A;Map position: 5
A;Introns: 326/3

Query Match   10.6%;   Score 75.5; DB 2; Length 402;
Best Local Similarity   24.3%;   Pred. No. 6.1;
Matches 33; Conservative 23; Mismatches 35; Indels 45; Gaps 8;

Qy   5 QQRVERFLQYF-----SAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLL---HC 56
           |||:|| |:::|| |:| | | | || || || | || ||: | |:|| ||
Db   202 EQRIDLFVKYFEDILSSVGHTIPVE-----ALYMQACNQNTVLRVLRHLDSRLLKTISI 255

Qy   57 RIEADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIV 116
          :|:| | ||::|:| :| :|: : |||||||||||||||||||||||||||||: |
Db   256 KIVE-DDQDKPCNWVLPLEIN--------------EIC------ELDQWKFAENI 289

Qy   117 SGFIEH--AAEVREY 129
          :|| || ||
Db   290 E--VEHLIDTDDVRHF 303

RESULT 7
S25365
CYC8 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: glucose repression mediator; protein YBR0908; protein YBR112c; SSN6 p
C;Species: Saccharomyces cerevisiae
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 11-Jan-2000
C;Accession: S25365; S48277; S45980; S25404; S25405; A30906; S44692
R;Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann, H.
Yeast 8, 397-408, 1992
A;Title: Molecular analysis of yeast chromosome II between CMD1 and LYS2: the excision r
A;Reference number: S25364; MUID:92327848
A;Accession: S25365
A;Molecule type: DNA
A;Residues: 1-966 <MAN>
A;Cross-references: EMBL:X66247; NID:g3548; PIDN:CAA46973.1; PID:g3550
R;Mannhaupt, G.; Stucka, R.; Ehnle, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A;Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A;Reference number: S48255; MUID:95208357
A;Accession: S48277
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-966 <MAW>
A;Cross-references: EMBL:X78993; NID:g476045; PIDN:CAA55615.1; PID:g476068
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1994
R;Feldmann, H.; Mannhaupt, G.; Schwarzlose, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A;Reference number: S45927
A;Accession: S45980
A;Molecule type: DNA

A;Residues: 1-966 <FE2>
A;Cross-references: EMBL:Z35981; NID:g536449; PIDN:CAA85069.1; PID:g536450; MIPS:YBR
R;Schultz, J.; Carlson, M.
Mol. Cell. Biol. 7, 3637-3645, 1987
A;Title: Molecular analysis of SSN6, a gene functionally related to the SNF1 protein
A;Reference number: S25404; MUID:88065502
A;Accession: S25404
A;Molecule type: DNA
A;Residues: 1-546,'K',548-966 <SCH>
A;Cross-references: EMBL:M17826; NID:g172725; PIDN:AAA35103.1; PID:g172726
R;Trumbly, R.J.
Gene 73, 97-111, 1988
A;Title: Cloning and characterization of the CYC8 gene mediating glucose repression
A;Reference number: S25405; MUID:89211964
A;Accession: S25405
A;Molecule type: DNA
A;Residues: 1-546,'K',548-966 <TRU>
A;Cross-references: EMBL:M23440; NID:g171349; PIDN:AAA34545.1; PID:g171350
C;Genetics:
A;Gene: SGD:CYC8; SSN6; CRT8
A;Cross-references: SGD:S0000316; MIPS:YBR112c
A;Map position: 2R
C;Function:
A;Description: required for complete derepression of ICL1; required for repression
C;Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repea
C;Keywords: nucleus; transcription regulation
F;224-257/Domain: tetratricopeptide repeat homology <TT1>
F;262-295/Domain: tetratricopeptide repeat homology <TT2>
F;296-329/Domain: tetratricopeptide repeat homology <TT3>
F;330-363/Domain: tetratricopeptide repeat homology <TT4>
F;365-398/Domain: tetratricopeptide repeat homology <TT5>

Query Match   10.6%;   Score 75.5; DB 2; Length 966;
Best Local Similarity   26.2%;   Pred. No. 17;
Matches 22; Conservative 15; Mismatches 34; Indels 13; Gaps 2;

Qy   64 QTSITLYSMLLQLNFEM---AAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGF 119
          |||| ||||||||| ||:|||||||||||||||||||||||||||||||||||
Db   97 QRAAELYERALLVNPELSDVWATLGHCYLMLDDLQRAYNAYQQALYHLSNPNVPKLWHGI 156

Qy   120 -----IEHAAEVREYIAQLD 134
             :|||||||||:|||||
Db   157 GILYDRYGSLDYAEEAFAKVLELD 180

RESULT 8
A34076
protein-tyrosine kinase (EC 2.7.1.112) receptor type eph 1 precursor - human
N;Alternate names: receptor tyrosine kinase eph
C;Species: Homo sapiens (man)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 04-Feb-2000
C;Accession: A34076; S44280
R;Hirai, H.; Maru, Y.; Hagiwara, K.; Nishida, J.; Takaku, F.
Science 238, 1717-1720, 1987
A;Title: A novel putative tyrosine kinase receptor encoded by the eph gene.
A;Reference number: A34076; MUID:88070650
A;Accession: A34076
A;Molecule type: mRNA
A;Residues: 1-984 <HIR>
A;Cross-references: GB:M18391; NID:g339716; PIDN:AAA36747.1; PID:g339717
A;Note: the sequence in GenBank entry HUMTKR, release 111.0, has the codons GCG for
R;Tuzi, N.L.
submitted to the EMBL Data Library, November 1993
A;Description: An EGFR/eph chimeric receptor possesses ligand stimulated tyrosine ki
A;Reference number: S44280
A;Accession: S44280
A;Molecule type: mRNA
A;Residues: 286-397,'A',399-580,'QRDRATDVDREDKLWLKPYVDLQAYEDPAQGALDF',583,625-984 <T
A;Cross-references: EMBL:Z27409; NID:g482916; PIDN:CAA81796.1; PID:g482917
C;Genetics:
A;Gene: GDB:EPHT1; EPH; EPHT
A;Cross-references: GDB:119875; OMIM:179610

A;Map position: 7q32-7q36
C;Superfamily: protein-tyrosine kinase, receptor type eph; fibronectin type III repeat H
C;Keywords: ATP; autophosphorylation; glycoprotein; kinase-related transforming protein;
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-984/Product: protein-tyrosine kinase receptor type eph 1 #status predicted <MAT>
F;548-568/Domain: transmembrane #status predicted <TMM>
F;630-895/Domain: protein kinase homology <KIN>
F;638-646/Region: protein kinase ATP-binding motif
F;918-984/Domain: SAM homology <SAM>
F;59,338,414,478/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   10.3%;   Score 73.5; DB 1; Length 984;
Best Local Similarity   27.5%;   Pred. No. 28;
Matches   36; Conservative   20; Mismatches   52; Indels   23; Gaps   6;

Qy   7 RVERFLQYFSAGCKT------PIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
          | :: || :: || |: |: |||:| : ||||||: ||
Db   94 RVHVELQFTVRDCKSFPGGAGPLGCKETFNLLYMESDQDVGI----QLRRPLFQKVTTVA 149

Qy   61 ADPQTSI-TLYSMLLQLNFEMAAM-------RGCWLALDE------LHNVRLCFQQSLEHLD 108
          || ||: ||:|||||||||||| ||||||| ||||||||||||| |
Db   150 ADQSFTIRDLASGSVKLNVERCSLGRLTRRGLYLAFHNPGACVALVSVRVFYQRCPETLN 209

Qy   109 E-ASFSDIVSG 118
          : :||:||:||
Db   210 GLAQFPDTLPG 220

RESULT 9
T31673
N-acetylglucosaminyltransferases (EC 2.4.1.-), chain p110 - rat
N;Alternate names: O-GlcNAc transferase
C;Species: Rattus norvegicus (Norway rat)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
C;Accession: T31673
R;Kreppel, L.K.; Blomberg, M.A.; Hart, G.W.
J. Biol. Chem. 272, 9308-9315, 1997
A;Title: Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characteri
A;Reference number: Z21056; MUID:97238869
A;Accession: T31673
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1036 <KRE>
A;Cross-references: EMBL:U76557; NID:g1931578; PID:g1931579; PIDN:AAC53121.1
A;Experimental source: strain Sprague Dawle; liver
C;Genetics:
A;Gene: OGT
C;Keywords: glycosyltransferase; hexosyltransferase; tandem repeat

Query Match   10.0%;   Score 71.5; DB 2; Length 1036;
Best Local Similarity   24.4%;   Pred. No. 47;
Matches   30; Conservative   21; Mismatches   51; Indels   21; Gaps   4;

Qy   24 HLKDGVCALYNEQDEEAAVLEV----------PQHSDSL--LLHCRIIEADPQTSITLY 70
          |||||||||||||||||||||| |||||| ||||||||||||: : ||
Db   281 HFPDAYCNLANALKEKGSVAEAEDCYNTALRLCPTHADSLNNLANIKREQGNIEEAVRLY 340

Qy   71 SMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEA----SFSDIVSGFIEHAAE 125
          :|||||||||||||||||||||||||||||||||||||:| :||||||||||||
Db   341 RKALEVFPEFAAAHS--NLASVLQQQGKLQEALMHYKEAIRISPIFADAYSNMGNTLKE 397

Qy   126 VRE 128
          :||
Db   398 MQD 400

RESULT 10
T26216
hypothetical protein W06A7.3c - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T26216
R;Ainscough, R.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z20173
A;Accession: T26216
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2484 <WIL>
A;Cross-references: EMBL:Z78066; PIDN:CAB51467.1; GSPDB:GN00023; CESP:W06A7
A;Experimental source: clone W06A7
C;Genetics:
A;Gene: CESP:W06A7.3c
A;Map position: 5
A;Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2339/2; 2463/2

Query Match   10.0%;   Score 71.5; DB 2; Length 2484;
Best Local Similarity   24.2%;   Pred. No. 1.4e+02;
Matches   31; Conservative   19; Mismatches   43; Indels   35; Gaps   6;

Qy   18 GCKTPIHLKDGVCALYNEQDEEA-------------------AVLEVPQHSDSLLLHCRIIEADPQT 65
          || || || || |||||||:::   :||: || : :||:| : | |
Db   1491 GCITDVDASD------VNEQDEESTLKILKVVPSEPSLLELDFTNDPKVIHVPIPLMEPAT 1545

Qy   66 SITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAE 125
          | | | ||| ||| | || ||| |: | : : |::
Db   1546 ------MYLE--EMVE----WIIADVKEV-----SEMEVVTESEISEMAPQVSESTCP 1587

Qy   126 VREYIAQL 133
          : | ::|
Db   1588 IPEPLADL 1595

RESULT 11
T26215
hypothetical protein W06A7.3a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26215
R;Ainscough, R.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z20173
A;Accession: T26215
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2607 <WIL>
A;Cross-references: EMBL:Z78066; PIDN:CAB01522.2; GSPDB:GN00023; CESP:W06A7.3a
A;Experimental source: clone W06A7
C;Genetics:
A;Gene: CESP:W06A7.3a
A;Map position: 5
A;Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2412/1; 2462/2; 2586/2

Query Match   10.0%;   Score 71.5; DB 2; Length 2607;
Best Local Similarity   24.2%;   Pred. No. 1.4e+02;
Matches   31; Conservative   19; Mismatches   43; Indels   35; Gaps   6;

Qy   18 GCKTPIHLKDGVCALYNEQDEEA-------------------AVLEVPQHSDSLLLHCRIIEADPQT 65
          || || || || |||||||:::   :||: || : :||:| : | |
Db   1491 GCITDVDASD------VNEQDEESTLKILKVVPSEPSLLELDFTNDPKVIHVPIPLMEPAT 1545

Qy   66 SITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAE 125
          | | | ||| ||| | || ||| |: | : : |::
Db   1546 ------MYLE--EMVE----WIIADVKEV-----SEMEVVTESEISEMAPQVSESTCP 1587

Qy   126 VREYIAQL 133
          : | ::|
Db   1588 IPEPLADL 1595

RESULT 12
T31653

hypothetical protein Y57A10A.ff - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T31653
R;Smye, R.
submitted to the EMBL Data Library, September 1999
A;Reference number: 221048
A;Accession: T31653
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1048 <WIL>
A;Cross-references: EMBL:AL117195; NID:e1549729; PIDN:CAB55036.1; CESP:Y57A10A.ff
A;Experimental source: clone Y57A10A
C;Genetics:
A;Gene: CESP:Y57A10A.ff
A;Introns: 34/1; 82/3; 156/3; 398/3; 556/3; 798/1; 863/3; 1024/1

Query Match   9.9%; Score 70.5; DB 2; Length 1048;
Best Local Similarity 36.7%; Pred. No. 61;
Matches   22; Conservative 10; Mismatches 25; Indels   3; Gaps   2;

Qy   21 TPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADP--QTSITLYSMLLQLNF 78
             ::| ||| :|: : | :| :|| : :||| :| |||:| |
Db   346 TPKHIECVSVTVRNESDKKHAESSFTSHGLSVLYFNRKMECDPAAKYDITLKTMVLQ-NF 404

RESULT 13
T23051
hypothetical protein H06A10.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T23051
R;Lennard, N.
submitted to the EMBL Data Library, October 1997
A;Reference number: 219663
A;Accession: T23051
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-291 <WIL>
A;Cross-references: EMBL:Z99773; PIDN:CAB16924.1; GSPDB:GN00028; CESP:H06A10.1
A;Experimental source: clone H06A10
C;Genetics:
A;Gene: CESP:H06A10.1
A;Map position: X
A;Introns: 16/1; 59/1; 94/2; 129/3; 189/1; 218/2
C;Superfamily: Caenorhabditis elegans hypothetical protein H06A10.1

Query Match   9.7%; Score 69.5; DB 2; Length 291;
Best Local Similarity 23.2%; Pred. No. 16;
Matches   29; Conservative 23; Mismatches 38; Indels 35; Gaps   6;

Qy   14 YFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITLYSML 73
          ||| | : || ||| : : :   :| :| : : :
Db   191 YFFAGAGSEIDQKSG--------KKAAIL---RSDQTLNYCPMLK-----DITDQDII 232

Qy   74 LQLNFEMAAMRGCWLALDELHNVRLCFQQSLE--HLDEASFSDIVSGFIEHAAEVREYIA 131
          ||: | :| || | : || | | ||
Db   233 IRLDQSQTIYDIEWISV------FCYKYSHDFGHLD------MGLVENEEQVPPYIP 277

Qy   132 QLDES 136
          |
Db   278 DISIS 282

RESULT 14
S76366
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S76366

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecho
s.
A;Reference number: S74322; MUID:97061201
A;Accession: S76366
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-393 <KAN>
A;Cross-references: EMBL:D64000; GB:AB001339; NID:g1001484; PIDN:BAA10218.1; PID:g
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: mannose-1-phosphate guanylyltransferase

Query Match   9.7%; Score 69.5; DB 2; Length 393;
Best Local Similarity 21.1%; Pred. No. 23;
Matches   31; Conservative 25; Mismatches 62; Indels 29; Gaps   4;

Qy   5 QQRVERFLQYFSAGCKTPIHLKDGV------CALYNEQDEEAAVLEVPQHSDSLL--- 53
              : :|| | | :| : :|:|| ||:| :|
Db   140 QDHLEQFLAHFQAKSLDSLILLRRVSNPSAFGVATVNDQGKVLALVEKPEHPPSNLALVG 199

Qy   54 ------LHCRIIEADPQ------TSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLC 99
                    :|: || | | | ||| | |:: :|
Db   200 LYFFAPTHQAIANIEPSARGELEITDAIQYLLISHDYRVESLQLKGWWLDTGKKDDLLAA 259

Qy   100 FQQSLEHLDEASFSDIVSGFIEHAAEV 126
          | :|||:|| ||| | | |:
Db   260 NQIILDTLVEKN----IQGTVDQSKI 282

RESULT 15
ZLNZMV
genome polyprotein - measles virus (strain Udem)
N;Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C;Species: measles virus
C;Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 11-Jun-1999
C;Accession: A28919
R;Blumberg, B.M.; Crowley, J.C.; Silverman, J.I.; Menonna, J.; Cook, S.D.; Dowling,
Virology 164, 487-497, 1988
A;Title: Measles virus L protein evidences elements of ancestral RNA polymerase.
A;Reference number: A28919; MUID:88219537
A;Accession: A28919
A;Molecule type: genomic RNA
A;Residues: 1-2183 <BLU>
A;Cross-references: GB:M20865; NID:g331776; PIDN:AAA46430.1; PID:g331777
C;Genetics:
A;Gene: L
C;Superfamily: parainfluenza virus RNA-directed RNA polymerase
C;Keywords: ATP; nucleotidyltransferase

Query Match   9.7%; Score 69.5; DB 1; Length 2183;
Best Local Similarity 20.0%; Pred. No. 1.8e+02;
Matches   38; Conservative 26; Mismatches 63; Indels 63; Gaps   7;

Qy   4 SQQRVERFLQYFSAGCK------------------TPIHLKDGVCALYNEQDEEAAVLEVP 46
          | || :| :|:| :| : |:| | :|| | :
Db   158 SSQWFEPFLFWFTVKTEMRSVIKSQTHTCHRRRHTPVFFTGSSVELLISRDLVAIISKES 217

Qy   47 QH----SDSLLLHCRIIEAD--PQTSITLYSMLLQLNFEMAAMRGCWLALD--------- 91
          :| | :|| :|| | :| || : | :| :| :
Db   218 QHVYLTFELVLMYCDVIEGRLMTETAMTIDARYTEL---LGRVRYMWKLIDGFFPALGN 274

Qy   92 ELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAE 125
          || :| | |::|||:|| ||| | | |:
Db   275 PTYQIVAMLEPLSLAYLQLRDITVELRGAFLNHCFTEIHDVLDQNGFSD--EGTYHELIE 332

Qy   126 VREYIAQLDE 135
          |:| || |
Db   333 ALDYIFITDD 342

us-09-596-784-4.rpr

Search completed: June 5, 2001, 18:18:12
Job time: 238 sec

```
                GenCore version 4.5
          Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:      June  5, 2001, 18:21:11 ; Search time 30.94 Seconds
                                       (without alignments)
                                       67.946 Million cell updates/sec

Title:          US-09-596-784-4
Perfect score:  715
Sequence:       1 MTSSQQRVERFLQYFSAGCK.........IEHAAEVREYIAQLDESSAA 139

Scoring table:  BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched:       296979 seqs, 15124130 residues

Total number of hits satisfying chosen parameters:      296979

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     Pending_Patents_AA.New:*
          1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
          2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
          3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
          4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
          5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
          6: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	Match	ID	Description
1	68	9.5	652	5		US-09-813-742-5	Sequence 5, Appli
2	68	9.5	1144	5		US-09-813-742-3	Sequence 3, Appli
3	66	9.2	466	5		US-09-739-449-12177	Sequence 12177, A
4	65	9.1	1275	1		PCT-US01-04098A-1583	Sequence 1583, Ap
5	65	9.1	1301	1		PCT-US01-04098A-3551	Sequence 3551, Ap
6	63.5	8.9	1867	5		US-09-824-574-5	Sequence 5, Appli
7	63	8.8	343	5		US-09-270-849B-181413	Sequence 181413,
8	62.5	8.7	947	6		US-60-248-505-702	Sequence 702, App
9	62.5	8.7	968	6		US-60-248-505-1139	Sequence 1139, Ap
10	62	8.7	511	1		PCT-US01-11988-2010	Sequence 2010, Ap
11	62	8.7	511	5		US-09-833-245-2010	Sequence 2010, Ap
12	61.5	8.6	313	5		US-09-270-849B-192462	Sequence 192462,
13	61	8.5	341	5		US-09-739-449-10007	Sequence 10007, A
14	61	8.5	528	6		US-60-248-505-758	Sequence 758, App
15	61	8.5	873	1		PCT-US00-35017A-1214	Sequence 1214, Ap
16	60.5	8.5	1120	1		PCT-US01-04098A-1515	Sequence 1515, Ap
17	60.5	8.5	1152	1		PCT-US01-04098A-3483	Sequence 3483, Ap
18	60	8.4	222	5		US-09-270-849B-194981	Sequence 194981,
19	60	8.4	839	1		PCT-US00-00897-7	Sequence 7, Appli
20	60	8.4	839	5		US-09-824-258-2	Sequence 2, Appli
21	60	8.4	839	5		US-09-824-258-8	Sequence 8, Appli
22	60	8.4	2263	1		PCT-US01-04098A-1662	Sequence 1662, Ap
23	59.5	8.3	536	1		PCT-US01-04098A-3539	Sequence 3539, Ap
24	59.5	8.3	1047	1		PCT-US01-04098A-1169	Sequence 1169, Ap
25	59.5	8.3	1047	1		PCT-US01-04098A-3137	Sequence 3137, Ap
26	59	8.3	149	5		US-09-270-849B-194537	Sequence 194537,
27	59	8.3	226	5		US-09-270-849B-182172	Sequence 182172,

28	59	8.3	366	5		US-09-270-849B-193802	Sequence 193802,
29	59	8.3	441	5		US-09-270-849B-183122	Sequence 183122,
30	59	8.3	1111	6		US-60-248-823-71	Sequence 71, Appl
31	58.5	8.2	107	5		US-09-640-211A-2165	Sequence 2165, Ap
32	58.5	8.2	370	5		US-09-345-236B-133	Sequence 133, App
33	58.5	8.2	557	5		US-09-739-449-12924	Sequence 12924, A
34	58.5	8.2	634	1		PCT-US01-04098A-1066	Sequence 1066, Ap
35	58.5	8.2	656	1		PCT-US01-04098A-3034	Sequence 3034, Ap
36	58	8.2	1226	5		US-09-739-449-9308	Sequence 9308, Ap
37	58	8.1	118	5		US-09-270-849B-194271	Sequence 194271,
38	58	8.1	641	5		US-09-768-877-26	Sequence 26, Appl
39	57.5	8.1	111	5		US-09-270-849B-186506	Sequence 186506,
40	57.5	8.0	281	5		US-09-270-849B-189982	Sequence 189982,
41	57.5	8.0	307	5		US-09-826-019-69	Sequence 69, Appl
42	57.5	8.0	615	1		PCT-US01-04098A-3300	Sequence 3300, Ap
43	57.5	8.0	622	1		PCT-US01-04098A-1332	Sequence 1332, Ap
44	57.5	8.0	951	1		PCT-US01-04098A-3596	Sequence 3596, Ap
45	57.5	8.0	1053	1		PCT-US01-04098A-1628	Sequence 1628, Ap

ALIGNMENTS

```
RESULT 1
US-09-813-742-5
;  Sequence 5, Application US/09813742
;  GENERAL INFORMATION:
;  APPLICANT:  Baker, Barbara
;  APPLICANT:  Dinesh-Kumar, S.P.
;  TITLE OF INVENTION: NON-PATHOGEN INDUCED SYSTEMIC ACQUIRED RESISTANCE (SAR) IN P
;  FILE REFERENCE: 042250/209601 (5830-12)
;  CURRENT APPLICATION NUMBER: US/09/813,742
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,027
;  PRIOR FILING DATE: 2000-03-21
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn version 3.0
;  SEQ ID NO 5
;     LENGTH: 652
;     TYPE: PRT
;     ORGANISM: Nicotiana glutinosa
US-09-813-742-5

Query Match             9.5%;  Score 68;  DB 5;  Length 652;
Best Local Similarity   23.6%;  Pred. No. 39;
Matches   38;  Conservative   23;  Mismatches   46;  Indels   54;  Gaps    8;

Qy      12 LQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITLYS 71
           |:|:|:|:|.|:|.:|:||.:.|:|      :.:|:::
Db     317 LDWFGNGSRIIITTRDK--HLIEKNDIIYEVTALPDHE--------SIQLFK 358

Qy      72 M------LLQLNFEMAAM------RGCWLAL---DELHNVRLC-FQQSLEHLDEASFSD 114
            :|:               ||| |||.||||.:.:||:.:::|:  :::|: |:|
Db     359 QHAFGKEVPNENFEKLSLEVVNYAKGLPLALKVWGSLLHNLRLTEWKSAIEHMKNNSYSG 418

Qy     115 IVSG---------FIEHA-----AEVREYIAQLDES 136
           |:                      |:
Db     419 IIDKLKISYDGLEPKQQEMFLDIACFLRGEEKDYILQILES 459

RESULT 2
US-09-813-742-3
;  Sequence 3, Application US/09813742
;  GENERAL INFORMATION:
;  APPLICANT:  Baker, Barbara
;  APPLICANT:  Dinesh-Kumar, S.P.
;  TITLE OF INVENTION: NON-PATHOGEN INDUCED SYSTEMIC ACQUIRED RESISTANCE (SAR) IN P
;  FILE REFERENCE: 042250/209601 (5830-12)
;  CURRENT APPLICATION NUMBER: US/09/813,742
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,027
;  PRIOR FILING DATE: 2000-03-21
```

```
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
;
; SEQ ID NO 3
; LENGTH: 1144
; TYPE: PRT
; ORGANISM: Nicotiana glutinosa
US-09-813-742-3

Query Match          9.5%;  Score 68;  DB 5;  Length 1144;
Best Local Similarity   23.6%;  Pred. No. 80;
Matches  38;  Conservative  23;  Mismatches  46;  Indels  54;  Gaps  8;

Qy     12 LQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITLYS 71
          |::||:|    | :::|::  :||:|   | :|:|
Db    317 LDWFGNGSRIIITTRDK--HLIEKNDIIYEVTALPDHE------------SIOLFK 358

Qy     72 M-----LLQLNFEMAAM------RGCWIAL----DELHNVRLC-FQQSLEHLDEASFSD 114
          |      |||:|||||      ||:||| ::   :::|| :|
Db    359 QHAFGKEVPNENFEKLSLEVVNYAKGLPLALKVWGSLLHNLRLTEWKSAIEHMKNNSYSG 418

Qy    115 IVSG--------FIEHA----AEVREYIAQLDES 136
          |:          |::| | :||
Db    419 IIDKLKISYDGLEPKQQEMFLDIACFLRGGEKKDYIIQILES 459

RESULT  3
US-09-739-449-12177
; Sequence 12177, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
;
; SEQ ID NO 12177
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-12177

Query Match          9.2%;  Score 66;  DB 5;  Length 466;
Best Local Similarity   25.2%;  Pred. No. 42;
Matches  31;  Conservative  19;  Mismatches  41;  Indels  32;  Gaps  6;

Qy     25 LKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITLYSMLLQLNFEMAAMR 84
          ||  |||  :|:  : |   |:  |:|::      :||:
Db    257 LKTPIAVLLN---EARTLE-PQHGDLVRAQADAMQAQVQSYLS-------RARIAAQR 303

Qy     85 GCWL--------ALDELHNVR-----------LCFQQ--SLEHLDEASFSDIVSGFIEHAAE 125
          |:||         :::| :|:              |  :||||:|||
Db    304 GSILARVEAEPALERLVRVMRRLNPDKQFVLNFEQPGAVLGMEQQDLEEVVGNLLENAAR 363

Qy    126 VRE 128
          | |
Db    364 FAE 366

RESULT  4
PCT-US01-04098A-1583
; Sequence 1583, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
;
; SEQ ID NO 1583
; LENGTH: 1275
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-1583

Query Match          9.1%;  Score 65;  DB 1;  Length 1275;
Best Local Similarity   25.9%;  Pred. No. 2e+02;
Matches  30;  Conservative  28;  Mismatches  46;  Indels  12;  Gaps  6;

Qy     30 CALYNEQDEEAAVLEVPQHSDSLLL-HC----RIIEADPQTSITLYSMLLQLNFEMAAMR 84
          |||||:|    :||:|||:|:  ||     :|   ::|| |:|:::|:|||
Db    581 CERTYQQHEAMKTQI--RESLLAKHALEKQQLFEAYERTHLQLRSELDKLNKEVTAVQ 637

Qy     85 GCWLAL-DELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQLDESSAA 139
          |:  ||  :|:||:|:    ::|:|      | ||::  ||| || |:: ||
Db    638 ECYLEVCREKDNLELTLRKTTEK-EQQTQEKIKEKLIQQLE--KEWQSKLDQTIKA 690

RESULT  5
PCT-US01-04098A-3551
; Sequence 3551, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
;
; SEQ ID NO 3551
; LENGTH: 1301
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
```

; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-270-849B-181413

```
Query Match               8.8%;  Score 63;  DB 5;  Length 343;
Best Local Similarity 18.4%;  Pred. No. 60;
Matches  26;  Conservative  27;  Mismatches  52;  Indels  36;  Gaps  4;

Qy    4  SQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRII---- 59
           : ||   :  |  |      :|   ::  |::  ||:| |  ::|:|:||
Db   59  SEQSLRRAL------KTVL------VYAESDSDLQDTSFPEQVQDLLFNLHMLLSDT 103

Qy   60  ------EADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFS 113
              ||:: :  |:|  : |  :  : | :
Db  104  VKMKEYQEDPEMLLDLMNRIAKGYQNNPDLRLTWLE--------NMAKKHRERANHT 152

Qy  114  DIVSGFIEHAAEVREYIAQLD 134
           | :::  |:|::|::|:
Db  153  EAAMCYVHAASLVSEYLSMLE 173
```

RESULT 8
US-60-248-505-702
; Sequence 702, Application US/60248505
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: c1000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 702
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Human
US-60-248-505-702

```
Query Match               8.7%;  Score 62.5;  DB 6;  Length 947;
Best Local Similarity 25.8%;  Pred. No. 2.5e+02;
Matches  34;  Conservative  21;  Mismatches  48;  Indels  29;  Gaps  7;

Qy   25  LKDGVCALYNEQDEEAAVLEVP-QHSDSLLLHC-RIIEADPQTSITLYSMLLQLNFEMA 81
           | :: | : :  |||  ||::  :  | |   :||  ||||  | |
Db  290  LKNDITAVTKEGKILLTNLEVPDTEGAVSSRLECHRQISGDWQTINKLLTQVHDMAFD-- 347

Qy   82  AMRGCWLALDELHNVRL-------CFQQSLEHL-----DEASFSDIVSGFIEHAAE 125
            |   |:  | |:|::          |: :        :  | :| |  |
Db  348  --GFW----EKHQLKMEQYLQLWKFEODFQQLVTEVEFLLNQQAELAD-VTGTIAQVKQ 399

Qy  126  VREYIAQLDESS 137
           ||:| |::|
Db  400  KIKKLENLDENS 411
```

RESULT 9
US-60-248-505-1139
; Sequence 1139, Application US/60248505
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
; TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: c1000918
; CURRENT APPLICATION NUMBER: US/60/248,505
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 1998
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1139
; LENGTH: 968

; NAME/KEY: misc_feature
; LOCATION: (1)..(1301)
; OTHER INFORMATION: Xaa = any amino acid or nothing
PCT-US01-04098A-3551

```
Query Match               9.1%;  Score 65;  DB 1;  Length 1301;
Best Local Similarity 25.9%;  Pred. No. 2e+02;
Matches  30;  Conservative  28;  Mismatches  46;  Indels  12;  Gaps  6;

Qy   30  CALYNEQDEEAAVLEVPQHSDSLLL-HC---RIIEADPQTSITLYSMLLQLNFEMAAMR 84
           |||        ::  |    ||||     ::|| ||||||  | | | ||
Db  607  CERTYQQHHEAMKTQI--RESLLAKHALEKQQLFEAYERTHLQLRSELDKLNKEVTAVQ 663

Qy   85  GCWLAL-DELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQLDESSAA 139
          | |||  ||  :|:|  :| :||  :: ||:| |::|:|  |:|:|:::|:
Db  664  ECYLEVCREKDNLELTLRKTTEK-EQQTQEKIKEKLIQQLE--KEWQSKLDQTIKA 716
```

RESULT 6
US-09-824-574-5
; Sequence 5, Application US/09824574
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Moilanen, Anu-Maarit
; APPLICANT: Palvimo, Jorma J.
; APPLICANT: Jnne, Olli A,
; TITLE OF INVENTION: ARIP4 Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 1867
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-824-574-5

```
Query Match               8.9%;  Score 63.5;  DB 5;  Length 1867;
Best Local Similarity 25.2%;  Pred. No. 4.7e+02;
Matches  37;  Conservative  27;  Mismatches  50;  Indels  33;  Gaps  9;

Qy   12  LQYFSAGCKTPIHLKDGV-----CALY-NEQDEE---AAVLEVPQHSDSLLLHCRII 59
           : ||   || |      |  |:  :: : :|||| |  | |||: ||
Db  430  IRYF-VSIKTNFLFAHGLLENVVRIVLYGLNQSDDDVQSVAASILTPITSEFVKLNNSTI 488

Qy   60  EADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQ--------SLEHLDEA 110
           || |||| | | ||    ||:: ::: :: ||: ||:
Db  489  EI---LVTTIWSLLARLDDISSSVGSIMDL---LAKLCDHQEVLDILKNKALEHPSEW 541

Qy  111  SFSDIVS---GFIEHA-AEVREYIAQL 133
          |  |:   ||:|| |:|:|:|
Db  542  SFKSLVPKLYPFLRHSISSVRRAVLNL 568
```

RESULT 7
US-09-270-849B-181413
; Sequence 181413, Application US/09270849B
; GENERAL INFORMATION:
; APPLICANT: Swimmer et al.
; TITLE OF INVENTION: Insect genome survey devices
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/270,849B
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 195450
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 181413
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

```
;    TYPE: PRT
; ORGANISM: Human
US-60-248-505-1139

Query Match    Best Local Similarity    8.7%;    Score 62.5;    DB 6;    Length 968;
Matches 34;    Conservative 21;    Mismatches 48;    Indels 29;    Gaps 7;

Qy    25 LKDGVCALYNEQDEEAAVLEVP--QHSDSLLLHC-RIIEADPQTSITLYSMLLQLNFEMA 81
         ||:.|:.||      ||||:.||    :.||:||| | :.||::
Db   300 LKNDITAVTKEGKILLTNLEVPDTEGAVSSRLECHRQISGDWQTINKLLTQVHDMAFD-- 357

Qy    82 AMRGCWLALDELHNVRL-------CFQQSLEHL--------DEASFSDIVSGFIEHAAE 125
Db   358 ---GFW----EKHQLKMEQYLQLWKFEQDFQQLVTEVEFLLNQQAELAD-VTGTIAQVKQ 409

Qy   126 VREYIAQLDESS 137
          : :||:|
Db   410 KIKKLENLDENS 421

RESULT 10
PCT-US01-11988-2010
; Sequence 2010, Application PC/TUS0111988
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: PCT/US01/11988
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR APPLICATION DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2010
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (171)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (358)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (388)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-11988-2010

Query Match    Best Local Similarity    8.7%;    Score 62;    DB 1;    Length 511;
Matches 23;    Conservative 29;    Mismatches 44;    Indels 12;    Gaps 5;

Qy    39 EAAVLEVPQHSDSLLLHCRIIEADPQTSITLYSMLLQLNFEMA------AMRGCWLALD 91
         :: . ||:.:| :|| |    .:|  |.:|      |::|          ||||:|
Db   195 EGSAKQLP-NAYFLLPKVQSIQLSQMQAETTNTSLLQVQLECSLHNKVCQQLKGCYLESD 253

Qy    92 ELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQLDESSAA 139
         :::: .||:|||. : ||     ||| ::| :|      || |
Db   254 DVLRLQMSIMVTMENFSK-EFEENTQHLLDHLSIV--YVA-TDESETS 297

RESULT 11
US-09-833-245-2010
; Sequence 2010, Application US/09833245
; GENERAL INFORMATION:
```

```
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCT
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229, 358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256, 931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199, 384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2010
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (171)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (358)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (388)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-2010

Query Match    Best Local Similarity    8.7%;    Score 62;    DB 5;    Length 511;
Matches 23;    Conservative 29;    Mismatches 44;    Indels 12;    Gaps 5;

Qy    39 EAAVLEVPQHSDSLLLHCRIIEADPQTSITLYSMLLQLNFEMA------AMRGCWLALD 91
         :: . ||:.:| :|| |    .:|  |.:|      |::|          ||||:|
Db   195 EGSAKQLP-NAYFLLPKVQSIQLSQMQAETTNTSLLQVQLECSLHNKVCQQLKGCYLESD 253

Qy    92 ELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQLDESSAA 139
         :::: .||:|||. : ||     ||| ::| :|      || |
Db   254 DVLRLQMSIMVTMENFSK-EFEENTQHLLDHLSIV--YVA-TDESETS 297

RESULT 12
US-09-270-849B-192462
; Sequence 192462, Application US/09270849B
; GENERAL INFORMATION:
; APPLICANT: Swimmer et al.
; TITLE OF INVENTION: Insect genome survey devices
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/270,849B
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 195450
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 192462
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-09-270-849B-192462

Query Match    Best Local Similarity    8.6%;    Score 61.5;    DB 5;    Length 313;
Matches 25;    Conservative 29;    Mismatches 29;    Indels 46;    Gaps 6;

Qy     7 RVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSL----LLHCRIIEAD 62
         | : | :|.     .:: |     ||  | : .:         ::::   :|::|: ||
Db   106 KAENYLKHMAA------VIEEYACSLEADSLRADKLKPAANHNDGIFRMVLVHKKLVVVD 159
Qy    63 PQTSITLYSMLLQLNFEMAAMR-------GCWLALDELHNVRLCFQQSLEHLDE----A 110
         |   :|| || ||:||                  :|| |::    || ||:
```

```
Db    160 P----------VNLGIAQMRMLNHMEKEAWMNIIVRHNQK-SFRDKFRYDKIYRDLQ 206
                      :||     :|| ||||
Qy    111 SFSDIVSGFIEHAAE 125

Db    207 KFENLTTKFIKIACE 221
          |:|    ||:|:||
Db    209 EVDVLE-SQLSESSCA 223
```

```
RESULT 13
US-09-739-449-10007
; Sequence 10007, Application US/09739449
; GENERAL INFORMATION:
;  APPLICANT: Hinkle, Gregory J.
;  APPLICANT: Slater, Steven C.
;  TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
;  FILE REFERENCE: 38-10(15490)C
;  CURRENT APPLICATION NUMBER: US/09/739,449
;  CURRENT FILING DATE: 2000-12-19
;  PRIOR APPLICATION NUMBER: US 09/514,000
;  PRIOR FILING DATE: 2000-02-23
;  NUMBER OF SEQ ID NOS: 13351
;  SEQ ID NO 10007
;  LENGTH: 341
;  TYPE: PRT
;  ORGANISM: Agrobacterium tumefaciens
US-09-739-449-10007
```

```
Query Match            8.5%;  Score 61;  DB 5;  Length 341;
Best Local Similarity   23.1%;  Pred. No. 98;
Matches   27;  Conservative  21;  Mismatches  43;  Indels  26;  Gaps  6:

Qy    25 LKDGVCALY------NEQDEAAVLEVPQHSDSLLLHCRIIEADPQTSITLYSMLLQL 76
            |:||:|||          |||: ||||| ::|||:|||:|||||  :|
Db    38 LGGGICAIYIPSGDLVLQEPDEQG------HYDTPL-AAPLERQPSLDIAMEKAAIAL 88

Qy    77 NFEMAAMRGCWLALDELHNVRLCFQQ---SLEHLD-EASFSDIVSGFIEHAAEVR 127
            :|:   ||::|| :| |         :||| | |:|||:||||||| :|
Db    89 RLDRA--GAWRLCRSTADIRKAFAEDIFAAVLHMEGCEAIGADLDALEVFYAAGLR 142
```

```
RESULT 14
US-60-248-505-758
; Sequence 758, Application US/60248505
; GENERAL INFORMATION:
;  APPLICANT: Beasley, Ellen
;  TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
;  TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
;  TITLE OF INVENTION: PROTEINS, AND USES THEREOF
;  FILE REFERENCE: c1000918
;  CURRENT APPLICATION NUMBER: US/60/248,505
;  CURRENT FILING DATE: 2000-11-15
;  NUMBER OF SEQ ID NOS: 1998
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 758
;  LENGTH: 528
;  TYPE: PRT
;  ORGANISM: Human
US-60-248-505-758
```

```
Query Match            8.5%;  Score 61;  DB 6;  Length 528;
Best Local Similarity   27.7%;  Pred. No. 1.7e+02;
Matches   38;  Conservative  17;  Mismatches  42;  Indels  40;  Gaps  9:

Qy    22 PITHLKDGVCALYNEQDEEAAVLEV-----------PQHSDSLLLHCRIIEADPQTSITL 69
            |:|.|:|| |||: |||:              |||||:|||||:|||||||||||
Db   108 PVH-EVGECARAIE-DKDNEVLEMSHMLESWNPSASPYENPGFNLLCLELEG-AQELVT- 163

Qy    70 YSMLLQLNFEMAAMRGCWLALDELH------NVRLCFQQSLEHLDEASFSDIVSGFIEH 122
            ::|||    ||::|| |:|||:|:       || |||||||:|||||:||||||||
Db   164 -------QLKAMGGVSVAGDLLHQLQSOVTNASLTL-KLLADSDQCSF-----GALQQ 208

Qy   123 AAEVREYIAQLDESSAA 139
```

```
RESULT 15
PCT-US00-35017A-1214
; Sequence 1214, Application PC/TUS0035017A
; GENERAL INFORMATION:
;  APPLICANT: Hyseq Inc
;  TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
;  FILE REFERENCE: 784PCT
;  CURRENT APPLICATION NUMBER: PCT/US00/35017A
;  CURRENT FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US09/488,725
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: US09/552,317
;  PRIOR FILING DATE: 2000-04-25
;  NUMBER OF SEQ ID NOS: 1478
;  SEQ ID NO 1214
;  LENGTH: 873
;  TYPE: PRT
;  ORGANISM: Homo sapiens
PCT-US00-35017A-1214
```

```
Query Match            8.5%;  Score 61;  DB 1;  Length 873;
Best Local Similarity   23.0%;  Pred. No. 3.3e+02;
Matches   29;  Conservative  19;  Mismatches  56;  Indels  22;  .Gaps  5:

Qy    20 KTPIHLKDGVCALYNEQDEEAA--VLEVPQHSDSLLLHCRIIEADPQTSITLYSMLLQLN 77
            |  :||  |||:    :|::|   ||:|||||:|||||||:|||||||||||||||
Db   513 KLVVHLYH-LCQMFPESASDAIKFVLRDAMHEMEMETKGRAALPGLDVLIYLKITGLL 571

Qy    78 FEMAAMRGCWLALDELHNV---RLCFQQSLEHLDEASFSDIVSG------FIEHAAEVR 127
            |:|   ||::|           :||||||:|||:|||:||:             |:||:
Db   572 FPTS--------DFWHPVVTPALVCLSQLLTKCPILSLQDVVKGLFVCCLFLEYVALSQ 622

Qy   128 EYIAQL 133
            :| :||
Db   623 RFIPEL 628
```

```
Search completed: June  5, 2001, 18:21:12
Job time: 388 sec
```

```
          GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        June  5, 2001, 18:20:30 ; Search time 139.63 Seconds
                                            (without alignments)
                                            160.129 Million cell updates/sec

Title:          US-09-596-784-4
Perfect score:  715
Sequence:       1 MTSSQQRVERFLQYFSAGCK........IEHAAEVREYIAQLDESSAA 139

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       1009251 seqs, 160854530 residues

Total number of hits satisfying chosen parameters:        1009251

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
               . Listing first 45 summaries

Database :   Pending_Patents_AA_Main:*
         1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
         2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
         3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
         4: /cgn2_6/ptodata/2/paa/US080_COMB.pep:*
         5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
         6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
         7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
         8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
         9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
        10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
        11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
        12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
        13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
        14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
        15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
        16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
        17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
        18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
        19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
        20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
        21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
        22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
        23: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match Length	DB	ID	Description
1	715	100.0	139	1	PCT-US98-15426-4	Sequence 4, Appli
2	715	100.0	139	15	US-09-120-663-4	Sequence 4, Appli
3	715	100.0	139	18	US-09-412-100-30	Sequence 30, Appl
4	715	100.0	139	18	US-09-431-614-10	Sequence 10, Appl
5	715	100.0	139	19	US-09-596-784-4	Sequence 4, Appli
6	85.5	12.0	1007	3	US-07-861-390D-7	Sequence 7, Appli
7	76.5	10.7	194	9	US-08-583-808-4	Sequence 4, Appli
8	76.5	10.7	1105	1	PCT-US96-14679A-2	Sequence 2, Appli
9	76.5	10.7	1105	1	PCT-US96-14679A-2	Sequence 2, Appli
10	76.5	10.7	1105	16	US-09-220-157-2	Sequence 2, Appli
11	76.5	10.7	1105	16	US-09-220-157A-2	Sequence 2, Appli
12	75	10.5	243	1	PCT-US97-07950-338	Sequence 338, App
13	75	10.5	243	12	US-08-858-207A-338	Sequence 338, App
14	75	10.5	317	1	US00-01132-2	Sequence 2, Appli
15	75	10.5	317	1	PCT-US00-17262-50	Sequence 50, App
16	75	10.5	317	1	PCT-US97-14436-492	Sequence 492, App
17	75	10.5	317	13	US-08-911-503-492	Sequence 492, App
18	75	10.5	317	13	US-08-911-503A-492	Sequence 492, App
19	75	10.5	317	16	US-09-238-477-2	Sequence 2, Appli
20	75	10.5	317	19	US-09-595-940-50	Sequence 50, Appl
21	75	10.5	2053	23	US-60-191-637-39778	Sequence 39778, A
22	75	10.5	2053	23	US-60-191-681-30805	Sequence 30805, A
23	74	10.3	589	19	US-09-540-236-3398	Sequence 3398, Ap
24	74	10.3	589	23	US-60-128-476-3291	Sequence 3291, Ap
25	73.5	10.3	181	23	US-60-170-374-3340	Sequence 3340, Ap
26	73.5	10.3	211	23	US-60-178-307-2751	Sequence 2751, Ap
27	73.5	10.3	984	8	US-08-449-645-19	Sequence 19, Appl
28	73.5	10.3	984	11	US-08-702-367-19	Sequence 19, Appl
29	73.5	10.3	984	17	US-09-378-759-19	Sequence 19, Appl
30	73	10.2	217	18	US-09-489-039A-7789	Sequence 7789, Ap
31	73	10.2	560	18	US-09-450-969-4699	Sequence 4699, Ap
32	73	10.2	662	23	US-60-207-422-133	Sequence 133, App
33	73	10.2	674	23	US-60-212-413-305	Sequence 305, App
34	73	10.2	674	23	US-60-229-518-252	Sequence 252, App
35	72.5	10.1	244	21	US-09-733-089-23131	Sequence 23131, A
36	71.5	10.0	425	16	US-09-248-796-14511	Sequence 14511, A
37	71.5	10.0	920	23	US-60-222-113-76	Sequence 76, Appl
38	70.5	9.9	103	15	US-09-107-532-5377	Sequence 5377, Ap
39	70.5	9.9	1445	21	US-09-733-089-22848	Sequence 22848, A
40	70.5	9.9	2183	16	US-09-269-367A-4	Sequence 4, Appli
41	70.5	9.9	2183	16	US-09-269-367A-12	Sequence 12, Appl
42	69.5	9.7	286	12	US-08-827-356-3003	Sequence 3003, Ap
43	69.5	9.7	286	20	US-09-611-529-3871	Sequence 3871, Ap
44	69.5	9.7	2183	16	US-09-269-367A-2	Sequence 2, Appli
45	69.5	9.7	2183	16	US-09-269-367A-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT  1
PCT-US98-15426-4
; Sequence 4, Application PC/TUS815426
; GENERAL INFORMATION:
;    APPLICANT: Cornell Research Foundation, Inc.
;    TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
;    TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
;    NUMBER OF SEQUENCES: 5
;    CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
;       STREET: P.O. Box 1051, Clinton Square
;       CITY: Rochester
;       STATE: New York
;       COUNTRY: U.S.A.
;       ZIP: 14603
;    COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: PCT/US98/15426
;       FILING DATE:
;       CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 60/055,105
;       FILING DATE: 06-AUG-1997
;    ATTORNEY/AGENT INFORMATION:
;       NAME: Goldman, Michael L.
;       REGISTRATION NUMBER: 30,727
;       REFERENCE/DOCKET NUMBER: 19603/1662
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (716) 263-1304
```

```
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO: 4:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  139 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
PCT-US98-15426-4

Query Match             100.0%;  Score 715;  DB 1;  Length 139;
Best Local Similarity   100.0%;  Pred. No. 2.8e-75;
Matches  139;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60

Qy     61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120

Qy    121 EHAAEVREYIAQLDESSAA 139
          |||||||||||||||||||
Db    121 EHAAEVREYIAQLDESSAA 139

RESULT 2
US-09-120-663-4
;  Sequence 4, Application US/09120663
;  GENERAL INFORMATION:
;      APPLICANT:  Bogdanove, Adam J.
;      APPLICANT:  Kim, Jihyun Francis
;      APPLICANT:  Wei, Zhong-Min
;      APPLICANT:  Beer, Steven V.
;      TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FROM
;      TITLE OF INVENTION:  ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
;      NUMBER OF SEQUENCES:  5
;  CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon, Hargrave, Devans & Doyle LLP
;      STREET:  P.O. Box 1051, Clinton Square
;      CITY:  Rochester
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  14603
;  COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/120,663
;      FILING DATE:
;      CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/055,105
;      FILING DATE:  06-AUG-1997
;  ATTORNEY/AGENT INFORMATION:
;      NAME:  Goldman, Michael L.
;      REGISTRATION NUMBER:  30,727
;      REFERENCE/DOCKET NUMBER:  19603/1661
;  TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (716) 263-1304
;      TELEFAX:  (716) 263-1600
;  INFORMATION FOR SEQ ID NO: 4:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  139 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-09-120-663-4
```

```
Query Match             100.0%;  Score 715;  DB 15;  Length 139;
Best Local Similarity   100.0%;  Pred. No. 2.8e-75;
Matches  139;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60

Qy     61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120

Qy    121 EHAAEVREYIAQLDESSAA 139
          |||||||||||||||||||
Db    121 EHAAEVREYIAQLDESSAA 139

RESULT 3
US-09-412-100-30
;  Sequence 30, Application US/09412100
;  GENERAL INFORMATION:
;      APPLICANT:  Wei, Zhong-Min
;      APPLICANT:  Fan, Hao
;      APPLICANT:  Niggemeyer, Jennifer L.
;      TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FRAGMENTS WHICH ARE
;      TITLE OF INVENTION:  ACTIVE BUT DO NOT ELICIT A HYPERSENSITIVE RESPONSE
;      FILE REFERENCE:  21829/31 (EBC-002)
;      CURRENT APPLICATION NUMBER:  US/09/412,100
;      CURRENT FILING DATE:  1999-10-04
;      EARLIER APPLICATION NUMBER:  60/103,050
;      EARLIER FILING DATE:  1998-10-05
;      NUMBER OF SEQ ID NOS:  39
;      SOFTWARE:  PatentIn Ver. 2.0
;  SEQ ID NO 30
;      LENGTH:  139
;      TYPE:  PRT
;      ORGANISM:  Erwinia amylovora
US-09-412-100-30
```

```
Query Match             100.0%;  Score 715;  DB 18;  Length 139;
Best Local Similarity   100.0%;  Pred. No. 2.8e-75;
Matches  139;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60

Qy     61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120

Qy    121 EHAAEVREYIAQLDESSAA 139
          |||||||||||||||||||
Db    121 EHAAEVREYIAQLDESSAA 139

RESULT 4
US-09-431-614-10
;  Sequence 10, Application US/09431614
;  GENERAL INFORMATION:
;      APPLICANT:  Wei, Zhong-Min
;      APPLICANT:  Schading, Richard L.
;      TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
;      TITLE OF INVENTION:  RESISTANCE
;      FILE REFERENCE:  21829/41 (EBC-003)
;      CURRENT APPLICATION NUMBER:  US/09/431,614
;      CURRENT FILING DATE:  1999-11-02
;      EARLIER APPLICATION NUMBER:  60/107,243
;      EARLIER FILING DATE:  1998-11-05
;      NUMBER OF SEQ ID NOS:  18
```

```
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 10
;  LENGTH: 139
;  TYPE: PRT
;  ORGANISM: Erwinia amylovora
US-09-431-614-10

Query Match             100.0%;  Score 715;  DB 18;  Length 139;
Best Local Similarity   100.0%;  Pred. No. 2.8e-75;
Matches 139;  Conservative   0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60

Qy   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120

Qy  121 EHAAEVREYIAQLDESSAA 139
         |||||||||||||||||||
Db  121 EHAAEVREYIAQLDESSAA 139

RESULT  5
US-09-596-784-4
;  Sequence 4, Application US/09596784
;  GENERAL INFORMATION:
;     APPLICANT: Bogdanove, Adam J.
;     APPLICANT: Kim, Jihyun Francis
;     APPLICANT: Wei, Zhong-Min
;     APPLICANT: Beer, Steven V.
;     TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
;     TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
;     NUMBER OF SEQUENCES: 5
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
;        STREET: P.O. Box 1051, Clinton Square
;        CITY: Rochester
;        STATE: New York
;        COUNTRY: U.S.A.
;        ZIP: 14603
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE: Floppy disk
;        COMPUTER: IBM PC compatible
;        OPERATING SYSTEM: PC-DOS/MS-DOS
;        SOFTWARE: PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: US/09/596,784
;        FILING DATE:
;        CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;        APPLICATION NUMBER: 09/120,663
;        FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;        NAME: Goldman, Michael L.
;        REGISTRATION NUMBER: 30,727
;        REFERENCE/DOCKET NUMBER: 19603/1661
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE: (716) 263-1304
;        TELEFAX: (716) 263-1600
;  INFORMATION FOR SEQ ID NO: 4:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH: 139 amino acids
;        TYPE: amino acid
;        STRANDEDNESS:
;        TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-09-596-784-4

Query Match             100.0%;  Score 715;  DB 19;  Length 139;
```

```
Best Local Similarity   100.0%;  Pred. No. 2.8e-75;
Matches 139;  Conservative   0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60

Qy   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120

Qy  121 EHAAEVREYIAQLDESSAA 139
         |||||||||||||||||||
Db  121 EHAAEVREYIAQLDESSAA 139

RESULT  6
US-07-861-390D-7
;  Sequence 7, Application US/07861390D
;  GENERAL INFORMATION:
;     APPLICANT: PAWSON et al.
;     TITLE OF INVENTION: METHOD FOR EXPRESSING A
;     TITLE OF INVENTION: PHOSPHORYLATED PROTEIN
;     NUMBER OF SEQUENCES: 8
;     CORRESPONDENCE ADDRESS:
;        ADDRESSEE: Linda M. Kurdydyk, Bereskin & Parr
;        STREET: 40 King Street West, P. O. Box 401
;        CITY: Toronto, Ontario
;        STATE: N/A
;        COUNTRY: Canada
;        ZIP: M5H 3Y2
;     COMPUTER READABLE FORM:
;        MEDIUM TYPE: Floppy disk
;        COMPUTER: IBM PC compatible
;        OPERATING SYSTEM: PC-DOS/MS-DOS
;        SOFTWARE: DOS Text File
;     CURRENT APPLICATION DATA:
;        APPLICATION NUMBER: US/07/861,390D
;        FILING DATE: 19920331
;        CLASSIFICATION: 435
;     ATTORNEY/AGENT INFORMATION:
;        NAME: J.G. Mullins
;        REGISTRATION NUMBER: 33073
;        REFERENCE/DOCKET NUMBER: 027-021
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE: (703) 684-1111
;  INFORMATION FOR SEQ ID NO: 7:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH: 1007 amino acids
;        TYPE: AMINO ACID
;        STRANDEDNESS: single
;        TOPOLOGY: linear
;     MOLECULE TYPE: protein
US-07-861-390D-7

Query Match              12.0%;  Score 85.5;  DB 3;  Length 1007;
Best Local Similarity    29.0%;  Pred. No. 1.1;
Matches 38;  Conservative 19;  Mismatches 19;  Indels 55;  Gaps  5;

Qy    7 RVERFLQYFSAGCKT-----PIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
          || ||:  ||    :     || |:|::  :
Db   95 RVHVELQFTVRDCKSPPGGAGPLGCKETFNLLYMESDQDVGXXXXIQLRRPLFQKVTIVA 154

Qy   61 ADPQTSI-TLYSMLLQLNFEMAAM------RGCWLALDE---------LHNVRLCFQQSLEHLD 108
          |  |:  :: ::|  |  || :       :: ::|
Db  155 ADQSFTIRDLASGSVKLNVERCSLGRLTRRGLYLAFHNPGACVALVSVRVFYQRCPETLN 214

Qy  109 E-ASFSDIVSG 118
          | |:  :  |
Db  215 GLAQPDILPG 225
```

RESULT 7
US-08-583-808-4
; Sequence 4, Application US/08583808
; GENERAL INFORMATION:
.; APPLICANT: Villeponteau, Bryant
.; APPLICANT: Feng, Junli
.; TITLE OF INVENTION: Identification of Telomerase-Related
.; TITLE OF INVENTION: Protein Components
.; NUMBER OF SEQUENCES: 42
.; CORRESPONDENCE ADDRESS:
.; ADDRESSEE: Townsend and Townsend and Crew LLP
.; STREET: Two Embarcadero Center, Eighth Floor
.; CITY: San Francisco
.; STATE: California
.; COUNTRY: USA
.; ZIP: 94111-3834
.; COMPUTER READABLE FORM:
.; MEDIUM TYPE: Floppy disk
.; COMPUTER: IBM PC compatible
.; OPERATING SYSTEM: PC-DOS/MS-DOS
.; SOFTWARE: PatentIn Release #1.0, Version #1.30
.; CURRENT APPLICATION DATA:
.; APPLICATION NUMBER: US/08/583,808
.; FILING DATE: 05-JAN-1996
.; CLASSIFICATION: 435
.; PRIOR APPLICATION DATA:
.; APPLICATION NUMBER: US 60/003,492
.; FILING DATE: 09-SEP-1995
.; ATTORNEY/AGENT INFORMATION:
.; NAME: Storella, John R.
.; REGISTRATION NUMBER: 32,944
.; REFERENCE/DOCKET NUMBER: 015389-001210US
.; TELECOMMUNICATION INFORMATION:
.; TELEPHONE: (415) 576-0200
.; TELEFAX: (415) 576-0300
US-08-583-808-4
.; INFORMATION FOR SEQ ID NO: 4:
.; SEQUENCE CHARACTERISTICS:
.; LENGTH: 194 amino acids
.; TYPE: amino acid
.; STRANDEDNESS:
.; TOPOLOGY: linear
.; MOLECULE TYPE: peptide
.; FEATURE:
.; NAME/KEY: Peptide
.; LOCATION: 1..194
.; OTHER INFORMATION: /note= "amino acids translated from
.; OTHER INFORMATION: nucleotide positions 1-582 of TPC2 cDNA
.; OTHER INFORMATION: (SEQ ID NO:3)"
US-08-583-808-4

Query Match   10.7%; Score 76.5; DB 9; Length 194;
Best Local Similarity 20.0%; Pred. No. 1.3;
Matches   25; Conservative   31; Mismatches   46; Indels   23; Gaps   3;

```
Qy   25 LKDGVCALYNEQDEEAAVLEV--------PQHSDSLLLHCRIIEADPQTSITLYSML 73
         |:|: || :|: :|||
Db   50 LEDKIRALKESKDQLESVLEVLHRQMEQYRDQPQHLEKIAYQQKLLQED-----LVHIR 103

Qy   74 LQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQL 133
         :|:| :| : | :: | : ::
Db  104 AELSRESTEMENAWNEYLKLENDVEQLKQTLQEQHRRAF------FFQEKSQIQKDLWRI 157

Qy  134 DESSA 138
         ::|
Db  158 EDVTA 162
```

RESULT 8
PCT-US96-14679-2
; Sequence 2, Application PC/TUS9614679
; GENERAL INFORMATION:
.; APPLICANT: VILLEPONTEAU, BRYANT
.; APPLICANT: FENG, JUNLI
.; APPLICANT: ANDREWS, WILLIAM H.
.; APPLICANT: ADAMS, ROBERT R.
.; TITLE OF INVENTION: METHODS AND REAGENTS FOR REGULATING
.; TITLE OF INVENTION: TELOMERE LENGTH AND TELOMERASE ACTIVITY
.; NUMBER OF SEQUENCES: 27
.; CORRESPONDENCE ADDRESS:
.; ADDRESSEE: GERON CORPORATION
.; STREET: 200 CONSTITUTION DRIVE
.; CITY: MENLO PARK
.; STATE: CALIFORNIA
.; COUNTRY: U.S.A.
.; ZIP: 94025
.; COMPUTER READABLE FORM:
.; MEDIUM TYPE: Floppy disk
.; COMPUTER: IBM PC compatible
.; OPERATING SYSTEM: PC-DOS/MS-DOS
.; SOFTWARE: PatentIn Release #1.0, Version #1.30
.; CURRENT APPLICATION DATA:
.; APPLICATION NUMBER: PCT/US96/14679
.; FILING DATE: 13-SEP-1996
.; CLASSIFICATION:
.; ATTORNEY/AGENT INFORMATION:
.; NAME: KASTER ESQ., KEVIN R.
.; REGISTRATION NUMBER: 32,704
.; REFERENCE/DOCKET NUMBER: 012/045PCT
.; TELECOMMUNICATION INFORMATION:
.; TELEPHONE: (415) 473-7779
.; TELEFAX: (415)473-7750
.; INFORMATION FOR SEQ ID NO: 2:
.; SEQUENCE CHARACTERISTICS:
.; LENGTH: 1105 amino acids
.; TYPE: amino acid
.; TOPOLOGY: linear
.; MOLECULE TYPE: protein
PCT-US96-14679-2

Query Match   10.7%; Score 76.5; DB 1; Length 1105;
Best Local Similarity 20.0%; Pred. No. 14;
Matches   25; Conservative   31; Mismatches   46; Indels   23; Gaps   3;

```
Qy   25 LKDGVCALYNEQDEEAAVLEV--------PQHSDSLLLHCRIIEADPQTSITLYSML 73
         |:|: || :|: :|||
Db  691 LEDKIRALKENKDQLESVLEVLHRQMEQYRDQPQHLEKIAYQQKLLQED-----LVHIR 744

Qy   74 LQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQL 133
         :|:| :| : | :: | : ::
Db  745 AELSRESTEMENAWNEYLKLENDVEQLKQTLQEQHRRAF------FFQEKSQIQKDLWRI 798

Qy  134 DESSA 138
         ::|
Db  799 EDVTA 803
```

RESULT 9
PCT-US96-14679A-2
; Sequence 2, Application PC/TUS9614679A
; GENERAL INFORMATION:
.; APPLICANT: VILLEPONTEAU, BRYANT
.; APPLICANT: FENG, JUNLI
.; APPLICANT: ANDREWS, WILLIAM H.
.; APPLICANT: ADAMS, ROBERT R.
.; TITLE OF INVENTION: METHODS AND REAGENTS FOR REGULATING
.; TITLE OF INVENTION: TELOMERE LENGTH AND TELOMERASE ACTIVITY
.; NUMBER OF SEQUENCES: 27
.; CORRESPONDENCE ADDRESS:
.; ADDRESSEE: GERON CORPORATION
.; STREET: 200 CONSTITUTION DRIVE
.; CITY: MENLO PARK
.; STATE: CALIFORNIA
.; COUNTRY: U.S.A.
.; ZIP: 94025

```
       COMPUTER READABLE FORM:
         MEDIUM TYPE: Floppy disk
         COMPUTER: IBM PC compatible
         OPERATING SYSTEM: PC-DOS/MS-DOS
         SOFTWARE: PatentIn Release #1.0, Version #1.30
       CURRENT APPLICATION DATA:
         APPLICATION NUMBER: PCT/US96/14679A
         FILING DATE: 13-SEP-1996
         CLASSIFICATION:
       ATTORNEY/AGENT INFORMATION:
         NAME: KASTER ESQ., KEVIN R.
         REGISTRATION NUMBER: 32,704
         REFERENCE/DOCKET NUMBER: 012/045PCT
       TELECOMMUNICATION INFORMATION:
         TELEPHONE: (415) 473-7779
         TELEFAX: (415)473-7750
     INFORMATION FOR SEQ ID NO: 2:
       SEQUENCE CHARACTERISTICS:
         LENGTH: 1105 amino acids
         TYPE: amino acid
         TOPOLOGY: linear
       MOLECULE TYPE: protein
PCT-US96-14679A-2

Query Match          10.7%;    Score 76.5;   DB 1;   Length 1105;
Best Local Similarity  20.0%;   Pred. No. 14;
Matches   25;   Conservative  31;   Mismatches  46;   Indels  23;   Gaps  3;

Qy    25  LKDGVCALYNEQDEEAAVLEV--------PQHSDSLLHCRIIEADPQTSITLYSML 73
             |:|:|| :||  ||| : :   :|||: :  :::|:|
Db   691  LEDKIRALKENKDQLESVLEVLHRQMEQYRDQPQHLEKIAYQQKLLQED-----LVHIR 744

Qy    74  LQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQL 133
             :|:|     |     |:|:  :|     :|:  :::|: :|:
Db   745  AELSRESTEMENAWNEYLKLENDVEQLKQTLQEQHRRAF-----FFQEKSQIQKDLWRI 798

Qy   134  DESSA 138
             ::|
Db   799  EDVTA 803

RESULT 10
US-09-220-157-2
; Sequence 2, Application US/09220157
; GENERAL INFORMATION:
;   APPLICANT: Villeponteau, Bryant
;   APPLICANT: Feng, Junli
;   APPLICANT: Andrews, William H.
;   APPLICANT: Adams, Robert R.
;   TITLE OF INVENTION: Methods and Reagents for Regulating
;   TITLE OF INVENTION: Telomere Length and Telomerase Activity
;   NUMBER OF SEQUENCES: 26
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Townsend and Townsend and Crew LLP
;   STREET: Two Embarcadero Center, Eighth Floor
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94111-3834
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/220,157
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/710,249
;   FILING DATE: 13-SEP-1996
;   APPLICATION NUMBER: US 08/583,808
```

```
;   FILING DATE: 05-JAN-1996
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/003,492
;   FILING DATE: 08-SEP-1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Storella, John R.
;   REGISTRATION NUMBER: 32,944
;   REFERENCE/DOCKET NUMBER: 015389-001220US
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (415) 576-0200
;   TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1105 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FEATURE:
;   NAME/KEY: Protein
;   LOCATION: 1..1105
;   OTHER INFORMATION:   /note= "deduced amino acid sequence of
;   OTHER INFORMATION:   TPC2 open reading frame"
US-09-220-157-2

Query Match          10.7%;    Score 76.5;   DB 16;   Length 1105;
Best Local Similarity  20.0%;   Pred. No. 14;
Matches   25;   Conservative  31;   Mismatches  46;   Indels  23;   Gaps  3;

Qy    25  LKDGVCALYNEQDEEAAVLEV--------PQHSDSLLHCRIIEADPQTSITLYSML 73
             |:|:|| :||  ||| : :   :|||: :  :::|:|
Db   691  LEDKIRALKENKDQLESVLEVLHRQMEQYRDQPQHLEKIAYQQKLLQED-----LVHIR 744

Qy    74  LQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQL 133
             :|:|     |     |:|:  :|     :|:  :::|: :|:
Db   745  AELSRESTEMENAWNEYLKLENDVEQLKQTLQEQHRRAF-----FFQEKSQIQKDLWRI 798

Qy   134  DESSA 138
             ::|
Db   799  EDVTA 803

RESULT 11
US-09-220-157A-2
; Sequence 2, Application US/09220157A
; GENERAL INFORMATION:
;   APPLICANT: Villeponteau, Bryant
;   APPLICANT: Feng, Junli
;   APPLICANT: Andrews, William H.
;   APPLICANT: Adams, Robert R.
;   TITLE OF INVENTION: Methods and Reagents for Regulating
;   TITLE OF INVENTION: Telomere Length and Telomerase Activity
;   NUMBER OF SEQUENCES: 26
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Townsend and Townsend and Crew LLP
;   STREET: Two Embarcadero Center, Eighth Floor
;   CITY: San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94111-3834
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/220,157A
;   FILING DATE:
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/08/710,249
;   FILING DATE: 13-SEP-1996
```

```
;     APPLICATION NUMBER: US 08/583,808
;     FILING DATE: 05-JAN-1996
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 60/003,492
;     FILING DATE: 08-SEP-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Storella, John R.
;     REGISTRATION NUMBER: 32,944
;     REFERENCE/DOCKET NUMBER: 015389-001220US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (415) 576-0200
;     TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1105 amino acids
;     TYPE: amino acid
;     STRANDEDNESS:
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FEATURE:
;     NAME/KEY: Protein
;     LOCATION: 1..1105
;   OTHER INFORMATION:   /note= "deduced amino acid sequence of
;   OTHER INFORMATION:   TPC2 open reading frame"
US-09-220-157A-2

Query Match             10.7%;   Score 76.5;  DB 16;  Length 1105;
Best Local Similarity   20.0%;   Pred. No. 14;
Matches   25;   Conservative   31; Mismatches  46; Indels  23; Gaps   3;

Qy   25 LKDGVCALYNEQDEEAAVLEV--------PQHSDSLLLHCRIIEADPQTSITLYSML 73
            :|:|:|| ::| :: :|||    :||| :|: :  || ||||
Db  691 LEDKIRALKENKDQLESVLEVLHRQMEQYRDQPQHLEKIAYQQKLLQED------LVHIR 744

Qy   74 LQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQL 133
            :|:.|  : |: : :|  ::::::::  :|: ::::::: .::
Db  745 AELSRESTEMENAWNEYLKLENDVEQLKQTLQEQHRRAF------FFQEKSQIQKDLWRI 798

Qy  134 DESSA 138
            ::    :|
Db  799 EDVTA 803

RESULT 12
PCT-US97-07950-338
; Sequence 338, Application PC/TUS9707950
;   GENERAL INFORMATION:
;     APPLICANT: Black, Michael
;     APPLICANT: Hodgson, John
;     APPLICANT: Knowles, David
;     APPLICANT: Nicholas, Richard
;     APPLICANT: Stodola, Robert
;     TITLE OF INVENTION: Novel Compounds
;     NUMBER OF SEQUENCES: 552
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: SmithKline Beecham Corporation
;     STREET: 709 Swedeland Road
;     CITY: King of Prussia
;     STATE: PA
;     COUNTRY: USA
;     ZIP: 19406-0939
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US97/07950
;     FILING DATE: 14-MAY-1997
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 60/017670
;     FILING DATE: 14-MAY-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Gimmi, Edward R
;     REGISTRATION NUMBER: 38,891
;     REFERENCE/DOCKET NUMBER: P50475
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 610-270-4478
;     TELEFAX: 610-270-5090
;     TELEX:
; INFORMATION FOR SEQ ID NO: 338:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 243 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: Protein
PCT-US97-07950-338

Query Match             10.5%;   Score 75;   DB 1;  Length 243;
Best Local Similarity   23.8%;   Pred. No. 2.6;
Matches   29;   Conservative   25; Mismatches  30; Indels  38; Gaps   7;

Qy   20 KTPIHLKDG--VCA----------------LYNEQDEEAAVLEVPQHSDSLLLHC 56
            | ||: :  :|                   :|: ::|:|:|| : :|||
Db  106 KKPISSRDGMKLCVETSTTFDDWVRQSEKDYQDMLIYLKENDFAKIGELTE-KNALAMH- 163

Qy   57 RIIEADPQTSITLYSMLLQLNFE-MAAMR------GCWLALDELHNVRL-CFQQSLEHL 107
            : ||: :| || ::  |: :  |     ||      | :|: :| || :  : |||
Db  164 ---ATTKTASPAFSYLTDASYEAMAFVRQLREKGEACYFTMDAGPNVKVFCQEKDLEHL 219

Qy  108 DE 109
            ||
Db  220 SE 221

RESULT 13
US-08-858-207A-338
; Sequence 338, Application US/08858207A
;   GENERAL INFORMATION:
;     APPLICANT: Black, Michael
;     APPLICANT: Hodgson, John
;     APPLICANT: Knowles, David
;     APPLICANT: Nicholas, Richard
;     APPLICANT: Stodola, Robert
;     TITLE OF INVENTION: Novel Compounds
;     NUMBER OF SEQUENCES: 552
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: SmithKline Beecham Corporation
;     STREET: 709 Swedeland Road
;     CITY: King of Prussia
;     STATE: PA
;     COUNTRY: USA
;     ZIP: 19406-0939
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/858,207A
;     FILING DATE: 09-MAY-1997
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 60/017670
;     FILING DATE: 14-MAY-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Gimmi, Edward R
;     REGISTRATION NUMBER: 38,891
;     REFERENCE/DOCKET NUMBER: P50475
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 610-270-4478
;     TELEFAX: 610-270-5090
```

```
;    TELEX:
; INFORMATION FOR SEQ ID NO:    338:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:   243 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  None
US-08-858-207A-338

Query Match             10.5%;   Score 75;   DB 12;   Length 243;
Best Local Similarity   23.8%;   Pred. No. 2.6;
Matches   29;  Conservative   25;  Mismatches   30;  Indels   38;  Gaps   7;

Qy     20 KTPIHLKDG--VCA-------------LYNEQDEEAAVLEVPQHSDSLLLHC 56
          | || :|| :|                :| :::: |: :|: ::| :| :|
Db    106 KKPISSRDGMKLCVETSTTFDDWVRQSEKDYQDMLIYLKENDFAKIGELTE-KNALAMH- 163

Qy     57 RIIEADPQTSITLYSMLLQLNFE-MAAMR-------GCWLALDELHNVRL-CFQQSLEHL 107
          :  |: :|| :|   :: | :| :::| :| |||
Db    164 ---ATTKTASPAFSYLTDASYEAMAFVRQLREKGEACYFTMDAGPNVKVFCQEKDLEHL 219

Qy    108 DE 109
          |
Db    220 SE 221
```

RESULT 14
PCT-US00-01132-2
; Sequence 2, Application PC/TUS0001132
; GENERAL INFORMATION:
;   APPLICANT: SmithKline Beecham Corporation
;   TITLE OF INVENTION: mvd
;   FILE REFERENCE: GM10188
;   CURRENT APPLICATION NUMBER: PCT/US00/01132
;   CURRENT FILING DATE: 2000-01-19
;   EARLIER APPLICATION NUMBER: US 09/238,477
;   EARLIER FILING DATE: 1999-01-28
;   NUMBER OF SEQ ID NOS: 2
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 317
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
PCT-US00-01132-2

```
Query Match             10.5%;   Score 75;   DB 1;   Length 317;
Best Local Similarity   23.8%;   Pred. No. 3.8;
Matches   29;  Conservative   25;  Mismatches   30;  Indels   38;  Gaps   7;

Qy     20 KTPIHLKDG--VCA-------------LYNEQDEEAAVLEVPQHSDSLLLHC 56
          | || :|| :|                :| :::: |: :|: ::| :| :|
Db    180 KKPISSRDGMKLCVETSTTFDDWVRQSEKDYQDMLIYLKENDFAKIGELTE-KNALAMH- 237

Qy     57 RIIEADPQTSITLYSMLLQLNFE-MAAMR-------GCWLALDELHNVRL-CFQQSLEHL 107
          :  |: :|| :|   :: | :| :::| :| |||
Db    238 ---ATTKTASPAFSYLTDASYEAMAFVRQLREKGEACYFTMDAGPNVKVFCQEKDLEHL 293

Qy    108 DE 109
          |
Db    294 SE 295
```

RESULT 15
PCT-US00-17262-50
; Sequence 50, Application PC/TUS0017262
; GENERAL INFORMATION:
;   APPLICANT: SMITHKLINE BEECHAM COPRORATION
;   APPLICANT: SMITHKLINE BEECHAM p.l.c.
;   TITLE OF INVENTION: MEVALONATE PATHWAY GENES
;   FILE REFERENCE: GM50062
;   CURRENT APPLICATION NUMBER: PCT/US00/17262
;   CURRENT FILING DATE: 2000-06-22
;   PRIOR APPLICATION NUMBER: US 60/140,519
;   PRIOR FILING DATE: 1999-06-22
;   PRIOR APPLICATION NUMBER: US 60/146,682
;   PRIOR FILING DATE: 1999-08-02
;   NUMBER OF SEQ ID NOS: 72
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 50
;   LENGTH: 317
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
PCT-US00-17262-50

```
Query Match             10.5%;   Score 75;   DB 1;   Length 317;
Best Local Similarity   23.8%;   Pred. No. 3.8;
Matches   29;  Conservative   25;  Mismatches   30;  Indels   38;  Gaps   7;

Qy     20 KTPIHLKDG--VCA-------------LYNEQDEEAAVLEVPQHSDSLLLHC 56
          | || :|| :|                :| :::: |: :|: ::| :| :|
Db    180 KKPISSRDGMKLCVETSTTFDDWVRQSEKDYQDMLIYLKENDFAKIGELTE-KNALAMH- 237

Qy     57 RIIEADPQTSITLYSMLLQLNFE-MAAMR-------GCWLALDELHNVRL-CFQQSLEHL 107
          :  |: :|| :|   :: | :| :::| :| |||
Db    238 ---ATTKTASPAFSYLTDASYEAMAFVRQLREKGEACYFTMDAGPNVKVFCQEKDLEHL 293

Qy    108 DE 109
          |
Db    294 SE 295


Search completed: June   5, 2001, 18:20:31
Job time: 357 sec
```

```
            GenCore version 4.5
    Copyright (c) 1993 - 2000  Compugen Ltd.

OM protein - protein search, using sw model

Run on:     June  5, 2001, 18:16:59 ; Search time 26.26 Seconds
                                      (without alignments)
                                      101.687 Million cell updates/sec

Title:          US-09-596-784-4
Perfect score:  715
Sequence:       1 MTSSQQRVERFLQYFSAGCK.........IEHAAEVREYIAQLDESSAA 139

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       185757 seqs, 1921857 residues

Total number of hits satisfying chosen parameters:      185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_AA:*
             1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
             2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
             3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
             4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
             5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
             6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	80.5	11.3	368	2	US-08-869-137-2	Sequence 2, Appli
2	76.5	10.7	1105	2	US-08-710-249-2	Sequence 2, Appli
3	73.5	10.3	984	2	US-08-673-789-9	Sequence 9, Appli
4	73.5	10.3	984	2	US-08-449-645A-19	Sequence 19, Appli
5	73.5	10.3	984	2	US-08-702-367A-19	Sequence 19, Appli
6	73.5	10.3	984	5	PCT-US95-04681-19	Sequence 19, Appl
7	69.5	9.7	2183	1	US-08-348-891A-7	Sequence 7, Appl
8	69.5	9.7	2183	2	US-08-905-817-7	Sequence 7, Appl
9	68.5	9.6	591	1	US-08-179-738-5	Sequence 5, Appl
10	68.5	9.6	591	1	US-08-179-738-10	Sequence 10, Appl
11	68.5	9.6	591	2	US-08-628-145-5	Sequence 5, Appl
12	68.5	9.6	591	2	US-08-628-145-10	Sequence 10, Appl
13	68.5	9.6	595	1	US-08-171-718-16	Sequence 16, Appl
14	68.5	9.6	595	3	US-08-478-087-16	Sequence 16, Appl
15	68.5	9.6	596	1	US-08-179-738-2	Sequence 2, Appl
16	68.5	9.6	596	1	US-08-179-738-3	Sequence 3, Appl
17	68.5	9.6	596	2	US-08-628-145-2	Sequence 2, Appl
18	68.5	9.6	596	2	US-08-628-145-3	Sequence 3, Appl
19	68	9.5	652	1	US-08-261-663A-6	Sequence 6, Appli
20	68	9.5	652	5	PCT-US95-07754A-6	Sequence 6, Appli
21	68	9.5	1143	2	US-08-310-912A-108	Sequence 108, App
22	68	9.5	1143	5	PCT-US95-04589-108	Sequence 108, App
23	68	9.5	1144	1	US-08-261-663A-2	Sequence 2, Appli
24	68	9.5	1144	1	US-08-261-663A-4	Sequence 4, Appli
25	68	9.5	1144	3	US-08-930-996A-9	Sequence 9, Appli
26	68	9.5	1144	5	PCT-US95-07754A-2	Sequence 2, Appli
27	68	9.5	1144	5	PCT-US95-07754A-4	Sequence 4, Appli
28	67.5	9.4	584	1	US-08-179-738-7	Sequence 7, Appli
29	67.5	9.4	584	2	US-08-628-145-7	Sequence 7, Appli
30	65.5	9.2	947	2	US-08-500-857A-4	Sequence 4, Appli
31	64.5	9.0	1151	1	US-08-286-889-37	Sequence 37, Appl
32	64.5	9.0	1151	1	US-08-485-618-37	Sequence 37, Appl
33	64.5	9.0	1151	1	US-08-362-652-37	Sequence 37, Appl
34	64.5	9.0	1151	2	US-08-605-672-37	Sequence 37, Appl
35	64.5	9.0	1151	2	US-08-482-293A-37	Sequence 37, Appl
36	64.5	9.0	1151	2	US-08-943-363-37	Sequence 37, Appl
37	64.5	9.0	1161	1	US-08-485-618-55	Sequence 55, Appl
38	64.5	9.0	1161	1	US-08-362-652-55	Sequence 55, Appl
39	64.5	9.0	1161	2	US-08-605-672-55	Sequence 55, Appl
40	64.5	9.0	1161	2	US-08-482-293A-55	Sequence 55, Appl
41	64.5	9.0	1161	2	US-08-943-363-55	Sequence 55, Appl
42	63.5	8.9	626	1	US-08-630-524-2	Sequence 2, Appli
43	63.5	8.9	626	1	US-07-938-782A-2	Sequence 2, Appli
44	63.5	8.9	626	2	PCT-US93-08131-2	Sequence 2, Appli
45	62.5	8.7	615	2	US-08-484-101B-38	Sequence 38, Appl

ALIGNMENTS

```
RESULT   1
US-08-869-137-2
; Sequence 2, Application US/08869137
; Patent No. 5856157
; GENERAL INFORMATION:
;   APPLICANT:  Schultz, David
;   APPLICANT:  Craig, Richard
;   APPLICANT:  Medford, June I.
;   APPLICANT:  Mumma, R.O.
;   APPLICANT:  Cox-Foster, Diana L.
;   TITLE OF INVENTION:  A No. 5856157e1   D9 14:0-ACP Fatty
;   TITLE OF INVENTION:  Acid Desaturase and Gene Therefor
;   NUMBER OF SEQUENCES:  2
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  The Webb Law Firm
;     STREET:  700 Koppers Building, 436 Seventh Avenue
;     CITY:  Pittsburgh
;     STATE:  PA
;     COUNTRY:  USA
;     ZIP:  15219-1818
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette
;     COMPUTER:  IBM Compatible
;     OPERATING SYSTEM:  DOS
;     SOFTWARE:  FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/869,137
;     FILING DATE:
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  60/018,957
;     FILING DATE:  04-JUN-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Johnson, Barbara E
;     REGISTRATION NUMBER:  31,198
;     REFERENCE/DOCKET NUMBER:  2034-970766
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  412-471-8815
;     TELEFAX:  412-471-4094
;     TELEX:
; INFORMATION FOR SEQ ID NO:  2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  368 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
;   FRAGMENT TYPE:  internal
US-08-869-137-2
```

Query Match   11.3%; Score 80.5;   DB 2;   Length 368;
Best Local Similarity 20.1%; Pred. No. 0.066;
Matches 37; Conservative 26; Mismatches 54; Indels 67; Gaps 8;

Qy   6 QRVERFLQYFSA-----GC-KTPIHLKDGVCALYNEQDEEAAVLE------VPQHSDSL 52
         ::!!:.!! | |!|!!|| .!!:!! | || !! |!!|
Db 169 RQVEKTIQYLIALGQDIGTEKNPYHL----FIYTSFQERATFISHANTAKLAQQHGDKQ 223

Qy 53 LLH-CRIIEADPQTSITLYSMLLQLNFEM-------------------------- 80
         | |!! | !! ::!!:!!
Db 224 LAQICGTIAADEKRHETAYTRIVDKLFELDPDETMSCLAHMKRKITMPAHLMRDGRDPH 283

Qy 81 ------AAMRGCWLALDELHNVRLCFQQSLEH-LDEASFSDIVSGFIEHAAEVREYIA 131
            !!: | | |! !! :|:!! | || :!:: |
Db 284 LFQHFSVVASRTGVYTVMDYIN------ILEHFVEKWNIEKITAGLSDKGREAQDYVC 335

Qy 132 QLDE 135
         :! |
Db 336 KLGE 339

RESULT 2
US-08-710-249-2
; Sequence 2, Application US/08710249
; Patent No. 5858777
; GENERAL INFORMATION:
; APPLICANT: Villeponteau, Bryant
; APPLICANT: Feng, Junli
; APPLICANT: Andrews, William H.
; APPLICANT: Adams, Robert R.
; TITLE OF INVENTION: Methods and Reagents for Regulating
; TITLE OF INVENTION: Telomere Length and Telomerase Activity
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,249
; FILING DATE: 13-SEP-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/583,808
; FILING DATE: 05-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/003,492
; FILING DATE: 08-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John R.
; REGISTRATION NUMBER: 32,944
; REFERENCE/DOCKET NUMBER: 015389-001220US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1105

; OTHER INFORMATION: /note= "deduced amino acid sequence of
; OTHER INFORMATION: TPC2 open reading frame"
US-08-710-249-2

Query Match   10.7%; Score 76.5;   DB 2;   Length 1105;
Best Local Similarity 20.0%; Pred. No. 1;
Matches 25; Conservative 31; Mismatches 46; Indels 23; Gaps 3;

Qy 25 LKDGVCALYNEQDEEAAVLEV----------PQHSDSLLLHCRIIEADPQTSITLYSML 73
         |:|:!| :!:.:!||! ||| .::!:: ||::!|
Db 691 LEDKIRALKENKDQLESVLEVLHRQMEQYRDQPQHLEKIAYQQKLLQED------LVHIR 744

Qy 74 LQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQL 133
         ::!: | | ::|: |
Db 745 AELSRESTEMENAWNEYLKLENDVEQLKQTLQEQHRRAF------FFQEKSQIQKDLWRI 798

Qy 134 DESSA 138
         :! |
Db 799 EDVTA 803

RESULT 3
US-08-673-789-9
; Sequence 9, Application US/08673789
; Patent No. 5814479
; GENERAL INFORMATION:
; APPLICANT: ZHOU, RENPING; SCHULZ, NICHOLAS,
; APPLICANT: T.; KROMER, LAWRENCE, F.; VANDE WOUDE,
; APPLICANT: GEORGE, F.
; TITLE OF INVENTION: BSK RECEPTOR LIKE
; TITLE OF INVENTION: TYROSINE KINASE AND LIGAND AND THEIR
; TITLE OF INVENTION: USE IN DIAGNOSTIC AND THERAPEUTIC
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/673,789
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/177,812
; FILING DATE: 04-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341
; REFERENCE/DOCKET NUMBER: 2026-4105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 984
; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
US-08-673-789-9

Query Match   10.3%; Score 73.5;   DB 2;   Length 984;
Best Local Similarity 27.5%; Pred. No. 2;

```
Matches   36;  Conservative   20;  Mismatches   52;  Indels   23;  Gaps   6;

Qy      7 RVERFLQYFSAGCKT-----PIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
          || ||:            ||:  ||:  ::|    :| :|:
Db     94 RVHVELQFTVRDCKSFPGGAGPLGCKETFNLLYMESDQDVGI----QLRRPLFQKVTTVA 149

Qy     61 ADPQTSI-TLYSMLLQLNFEMAAM-----RGCWLALDE-----LHNVRLCFQQSLEHLD 108
          ||  ||  |::|||  ::          :|         ::|:  :|:
Db    150 ADQSFTIRDLASGSVKLNVERCSLGRLTRRGLYLAFHNPGACVALVSVRVFYQRCPETLN 209

Qy    109 E-ASFSDIVSG 118
          |:   |::|
Db    210 GLAQFPDTLPG 220

RESULT 4
US-08-449-645A-19
; Sequence 19, Application US/08449645A
; Patent No. 5981245
; GENERAL INFORMATION:
;   APPLICANT:  Fox, Gary M.
;   TITLE OF INVENTION:  EPH-Like Receptor Protein Tyrosine
;   TITLE OF INVENTION:  Kinases
;   NUMBER OF SEQUENCES:  43
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Amgen Patent Operations/RBW
;     STREET:  1840 Dehavilland Drive
;     CITY:  Thousand Oaks
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  91320
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/449,645A
;     FILING DATE:
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Winter, Robert B.
;     REFERENCE/DOCKET NUMBER:  A-287
; INFORMATION FOR SEQ ID NO:  19:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  984 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-449-645A-19

Query Match   10.3%;  Score 73.5;  DB 2;  Length 984;
Best Local Similarity   27.5%;  Pred. No.2;
Matches   36;  Conservative   20;  Mismatches   52;  Indels   23;  Gaps   6;

Qy      7 RVERFLQYFSAGCKT-----PIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
          || ||:            ||:  ||:  ::|    :| :|:
Db     94 RVHVELQFTVRDCKSFPGGAGPLGCKETFNLLYMESDQDVGI----QLRRPLFQKVTTVA 149

Qy     61 ADPQTSI-TLYSMLLQLNFEMAAM-----RGCWLALDE-----LHNVRLCFQQSLEHLD 108
          ||  ||  |::|||  ::          :|         ::|:  :|:
Db    150 ADQSFTIRDLASGSVKLNVERCSLGRLTRRGLYLAFHNPGACVALVSVRVFYQRCPETLN 209

Qy    109 E-ASFSDIVSG 118
          |:   |::|
Db    210 GLAQFPDTLPG 220

RESULT 5
US-08-702-367A-19
; Sequence 19, Application US/08702367A
; Patent No. 5981246
; GENERAL INFORMATION:
;   APPLICANT:  Fox, Gary M.
;   TITLE OF INVENTION:  EPH-Like Receptor Protein Tyrosine
;   TITLE OF INVENTION:  Kinases
;   NUMBER OF SEQUENCES:  43
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Amgen Patent Operations/RBW
;     STREET:  1840 Dehavilland Drive
;     CITY:  Thousand Oaks
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  91320
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/702,367A
;     FILING DATE:
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Winter, Robert B.
;     REFERENCE/DOCKET NUMBER:  A-287
; INFORMATION FOR SEQ ID NO:  19:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  984 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-08-702-367A-19

Query Match   10.3%;  Score 73.5;  DB 2;  Length 984;
Rest Local Similarity   27.5%;  Pred. No. 2;
Matches   36;  Conservative   20;  Mismatches   52;  Indels   23;  Gaps   6;

Qy      7 RVERFLQYFSAGCKT-----PIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
          || ||:            ||:  ||:  ::|    :| :|:
Db     94 RVHVELQFTVRDCKSFPGGAGPLGCKETFNLLYMESDQDVGI----QLRRPLFQKVTTVA 149

Qy     61 ADPQTSI-TLYSMLLQLNFEMAAM-----RGCWLALDE-----LHNVRLCFQQSLEHLD 108
          ||  ||  |::|||  ::          :|         ::|:  :|:
Db    150 ADQSFTIRDLASGSVKLNVERCSLGRLTRRGLYLAFHNPGACVALVSVRVFYQRCPETLN 209

Qy    109 E-ASFSDIVSG 118
          |:   |::|
Db    210 GLAQFPDTLPG 220

RESULT 6
PCT-US95-04681-19
; Sequence 19, Application PC/TUS9504681
; GENERAL INFORMATION:
;   APPLICANT:  Fox, Gary M.
;   TITLE OF INVENTION:  EPH-Like Receptor Protein Tyrosine
;   TITLE OF INVENTION:  Kinases
;   NUMBER OF SEQUENCES:  28
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Amgen Patent Operations/RBW
;     STREET:  1840 Dehavilland Drive
;     CITY:  Thousand Oaks
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  91320
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
```

```
;  ;       APPLICATION NUMBER:   PCT/US95/04681
;  ;       FILING DATE:
;  ;       CLASSIFICATION:
;  ;       ATTORNEY/AGENT INFORMATION:
;  ;       NAME: Winter, Robert B.
;  ;       REFERENCE/DOCKET NUMBER:  A-287
;  INFORMATION FOR SEQ ID NO: 19:
;  ;       SEQUENCE CHARACTERISTICS:
;  ;       LENGTH: 984 amino acids
;  ;       TYPE: amino acid
;  ;       STRANDEDNESS: single
;  ;       TOPOLOGY: linear
;  ;       MOLECULE TYPE: protein
PCT-US95-04681-19

Query Match        10.3%;   Score 73.5;   DB 5;   Length 984;
Best Local Similarity  27.5%;   Pred. No. 2;   Indels 23;   Gaps 6;
Matches 36;  Conservative 20;  Mismatches 52;  Indels 23;  Gaps 6;

Qy    7  RVERFLQYFSAGCKT------PIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLHCRIIE 60
              : :: :  :||       | ||||| ||||  | :|||  | :||  |  || :
Db   94  RVHVELQFTVRDCKSFPGGAGPLGCKETFNLLYMESDQDVGI----QLRRPLFQKVTTVA 149

Qy   61  ADPQTSI-TLYSMLLQLNFEMAAM-----RGCWLALDE-----LHNVRLCFQQSLEHLD 108
              |  ||  | ::|||||  ||        | :|| ||        ::| ||| ::|:|
Db  150  ADQSFTIRDLASGSVKLNVERCSLGRLTRRGLYLAFHNPGACVALVSVRVFYQRCPETLN 209

Qy  109  E-ASFSDIVSG 118
             | | :|:|
Db  210  GLAQFPDTLPG 220

RESULT 7
US-08-348-891A-7
;  Sequence 7, Application US/08348891A
;  Patent No. 5564136
;  GENERAL INFORMATION:
;  ;       APPLICANT: SASAKI, Keiko
;  ;       APPLICANT: MORI, Takayuki
;  ;       APPLICANT: MAKINO, Satoshi
;  ;       TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
;  ;       TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
;  ;       TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
;  ;       NUMBER OF SEQUENCES: 19
;  ;       CORRESPONDENCE ADDRESS:
;  ;       ADDRESSEE: YOUNG & THOMPSON
;  ;       STREET: 745 South 23rd Street
;  ;       CITY: Arlington
;  ;       STATE: Virginia
;  ;       COUNTRY: USA
;  ;       ZIP: 22202
;  COMPUTER READABLE FORM:
;  ;       MEDIUM TYPE: Floppy disk
;  ;       COMPUTER: IBM PC compatible
;  ;       OPERATING SYSTEM: PC-DOS/MS-DOS
;  ;       SOFTWARE: PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;  ;       APPLICATION NUMBER: US/08/348,891A
;  ;       FILING DATE: 25-NOV-1994
;  ;       CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;  ;       APPLICATION NUMBER: US 07/848,400
;  ;       FILING DATE: 10-MAR-1992
;  PRIOR APPLICATION DATA:
;  ;       APPLICATION NUMBER: JP 3-293625
;  ;       FILING DATE: 14-OCT-1991
;  ;       ATTORNEY/AGENT INFORMATION:
;  ;       NAME: PATCH, Andrew J.
;  ;       REGISTRATION NUMBER: 32,925
;  ;       REFERENCE/DOCKET NUMBER: KP-7501
;  ;       TELECOMMUNICATION INFORMATION:
;  ;       TELEPHONE: 703-521-2297
```

```
;  ;       TELEFAX: 703-685-0573
;  ;       TELEX: 248425 EMBON
;  INFORMATION FOR SEQ ID NO: 7:
;  ;       SEQUENCE CHARACTERISTICS:
;  ;       LENGTH: 2183 amino acids
;  ;       TYPE: amino acid
;  ;       TOPOLOGY: linear
;  ;       MOLECULE TYPE: protein
US-08-348-891A-7

Query Match         9.7%;   Score 69.5;   DB 1;   Length 2183;
Best Local Similarity  20.0%;   Pred. No. 20;   Indels 63;   Gaps 7;
Matches 38;  Conservative 26;  Mismatches 63;  Indels 63;  Gaps 7;

Qy    4  SQQRVERFLQYFSAGCK-----------------TPIHLKDGVCALYNEQDEEAAVLEVP 46
             | | :|||: : ||:       ||:          |||  :||
Db  158  SSQWFEPFLFWFTVKTEMRSVIKSQTHTCHRRRHTPVFFTGSSVELLISRDLVAIISKES 217

Qy   47  QH------SDSLLHCRIIEAD--PQTSITLYSMLLQLNFEMAAMRGCWLALD-------- 91
             :|      :|:  |:| | |    |: |:| :|  |  |  |:   |::
Db  218  QHVYLTFELVLMYCDVIEGRLMTETAMTIDARYTEL---LGRVRYMWKLIDGFFPALGN 274

Qy   92  -----------------------------ELHNVRL-CFQQSLEHLDEASFSDIVSGFIEHAAE 125
                                         :|||||  ||| || | |||  || ||
Db      275  PTYQIVAMLEPLSLAYLQLRDITVELRGAFLNHCFTEIHDVLDQNGFSD--EGTYHELIE 332

Qy  126  VREYIAQLDE 135
             ||:  ||:
Db  333  ALDYIFITDD 342

RESULT 8
US-08-905-817-7
;  Sequence 7, Application US/08905817
;  Patent No. 5824777
;  GENERAL INFORMATION:
;  ;       APPLICANT: SASAKI, Keiko
;  ;       APPLICANT: MORI, Takayuki
;  ;       APPLICANT: MAKINO, Satoshi
;  ;       TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
;  ;       TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
;  ;       TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
;  ;       NUMBER OF SEQUENCES: 19
;  ;       CORRESPONDENCE ADDRESS:
;  ;       ADDRESSEE: YOUNG & THOMPSON
;  ;       STREET: 745 South 23rd Street
;  ;       CITY: Arlington
;  ;       STATE: Virginia
;  ;       COUNTRY: USA
;  ;       ZIP: 22202
;  COMPUTER READABLE FORM:
;  ;       MEDIUM TYPE: Floppy disk
;  ;       COMPUTER: IBM PC compatible
;  ;       OPERATING SYSTEM: PC-DOS/MS-DOS
;  ;       SOFTWARE: PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;  ;       APPLICATION NUMBER: US/08/905,817
;  ;       FILING DATE: 04-AUG-1997
;  ;       CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;  ;       APPLICATION NUMBER: US 08/348,891
;  ;       FILING DATE: 25-NOV-1994
;  PRIOR APPLICATION DATA:
;  ;       APPLICATION NUMBER: US 07/848,400
;  ;       FILING DATE: 10-MAR-1992
;  PRIOR APPLICATION DATA:
;  ;       APPLICATION NUMBER: JP 3-293625
;  ;       FILING DATE: 14-OCT-1991
;  ;       ATTORNEY/AGENT INFORMATION:
;  ;       NAME: PATCH, Andrew J.
;  ;       REGISTRATION NUMBER: 32,925
;  ;       REFERENCE/DOCKET NUMBER: KP-7501A
```

```
;      TELECOMMUNICATION INFORMATION:
;         TELEPHONE: 703-521-2297
;         TELEFAX: 703-685-0573
;         TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 7:
;         SEQUENCE CHARACTERISTICS:
;         LENGTH: 2183 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;         MOLECULE TYPE: protein
US-08-905-817-7

Query Match      9.7%;  Score 69.5; DB 2; Length 2183;
Best Local Similarity  20.0%; Pred. No.20;
Matches  38; Conservative 26; Mismatches  63; Indels  63; Gaps  7;

Qy    4 SQQRVERFLQYFSAGCK---------------------TPIHLKDGVCALYNEQDEEAAVLEVP 46
          :||    ||  : ||                   ||| ||  :   ||  |||||||||
Db  158 SSQWFEPFLFWFTVKTEMRSVIKSQTHTCHRRRHTPVFFTGSSVELLISRDLVAIISKES 217

Qy   47 QH------SDSLLHCRIIEAD--PQTSITLYSMLLQLNFEMAAMRGCWLALD--------- 91
         ||       :|:|:|:||    |:|::||   :  | :|    :|  :|
Db  218 QHVYYLTFELVLMYCDVIEGRLMTETAMTIDARYTEL--LGRVRYMWKLIDGFFPALGN 274

Qy   92 -------------------ELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAE 125
                            :|||||  |||  |||  :||   ||
Db  275 PTYQIVAMLEPLSLAYLQLRDITVELRGAFLNHCFTEIHDVLDQNGFSD--EGTYHELIE 332

Qy  126 VREYIAQLDE 135
          :|| ||
Db  333 ALDYIFITDD 342

RESULT 9
US-08-179-738-5
; Sequence 5, Application US/08179738
; Patent No. 5578462
; GENERAL INFORMATION:
;         APPLICANT: Seizinger, Bernd R.
;         APPLICANT: Kley, Nikolai A.
;         APPLICANT: Bianchi, Albert B.
;         TITLE OF INVENTION: No. 5578462el NF2 Isoforms
;         NUMBER OF SEQUENCES: 26
;         CORRESPONDENCE ADDRESS:
;         ADDRESSEE: Reed & Robins
;         STREET: 635 Bryant Street
;         CITY: Palo Alto
;         STATE: California
;         COUNTRY: U.S.A
;         ZIP: 94301
;         COMPUTER READABLE FORM:
;         MEDIUM TYPE: Floppy disk
;         COMPUTER: IBM PC compatible
;         OPERATING SYSTEM: PC-DOS/MS-DOS
;         SOFTWARE: PatentIn Release #1.0, Version #1.25
;         CURRENT APPLICATION DATA:
;         APPLICATION NUMBER: US/08/179,738
;         FILING DATE: 10-JAN-1994
;         CLASSIFICATION: 530
;         ATTORNEY/AGENT INFORMATION:
;         NAME: Robins, Roberta L.
;         REGISTRATION NUMBER: 33,208
;         REFERENCE/DOCKET NUMBER: 5998-0017
;         TELECOMMUNICATION INFORMATION:
;         TELEPHONE: (415) 617-8999
;         TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 5:
;         SEQUENCE CHARACTERISTICS:
;         LENGTH: 591 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;         MOLECULE TYPE: protein
```

```
;         FRAGMENT TYPE: N-terminal
;         ORIGINAL SOURCE:
;             ORGANISM: murine
US-08-179-738-5

Query Match      9.6%;  Score 68.5; DB 1; Length 591;
Best Local Similarity  25.7%; Pred. No. 4;
Matches  27; Conservative 22; Mismatches  43; Indels  13; Gaps  4;

Qy   33 YNEQDEEAAVLEVPQHSDSLLHCRIIE---ADPQTSITLYSMLLQL--NFEMAAMRG 85
          :|  ||  ||  |||      :|:      ||:  :|::||    |:  |    ::|:: |
Db  101 YPENAEELVQEITQHLFFLQVKKQILDEKVYCPPEASVLLASYAVQAKYGDYDPSVHKR 160

Qy   86 CWLALDELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAEVRE 128
         |:|:  :|||   |  ||  :||    ||    ||   ||
Db  161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201

RESULT 10
US-08-179-738-10
; Sequence 10, Application US/08179738
; Patent No. 5578462
; GENERAL INFORMATION:
;         APPLICANT: Seizinger, Bernd R.
;         APPLICANT: Kley, Nikolai A.
;         APPLICANT: Bianchi, Albert B.
;         TITLE OF INVENTION: No. 5578462el NF2 Isoforms
;         NUMBER OF SEQUENCES: 26
;         CORRESPONDENCE ADDRESS:
;         ADDRESSEE: Reed & Robins
;         STREET: 635 Bryant Street
;         CITY: Palo Alto
;         STATE: California
;         COUNTRY: U.S.A
;         ZIP: 94301
;         COMPUTER READABLE FORM:
;         MEDIUM TYPE: Floppy disk
;         COMPUTER: IBM PC compatible
;         OPERATING SYSTEM: PC-DOS/MS-DOS
;         SOFTWARE: PatentIn Release #1.0, Version #1.25
;         CURRENT APPLICATION DATA:
;         APPLICATION NUMBER: US/08/179,738
;         FILING DATE: 10-JAN-1994
;         CLASSIFICATION: 530
;         ATTORNEY/AGENT INFORMATION:
;         NAME: Robins, Roberta L.
;         REGISTRATION NUMBER: 33,208
;         REFERENCE/DOCKET NUMBER: 5998-0017
;         TELECOMMUNICATION INFORMATION:
;         TELEPHONE: (415) 617-8999
;         TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 10:
;         SEQUENCE CHARACTERISTICS:
;         LENGTH: 591 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;         MOLECULE TYPE: protein
;         FRAGMENT TYPE: N-terminal
;         ORIGINAL SOURCE:
;             ORGANISM: Homo sapien
US-08-179-738-10

Query Match      9.6%;  Score 68.5; Pred. No. 4;
Best Local Similarity  25.7%;
Matches  27; Conservative 22; Mismatches  43; Indels  13; Gaps  4;

Qy   33 YNEQDEEAAVLEVPQHSDSLLHCRIIE---ADPQTSITLYSMLLQL--NFEMAAMRG 85
          :|  ||  ||  |||      :|:      ||:  :|::||    |:  |    ::|:: |
Db  101 YPENAEELVQEITQHLFFLQVKKQILDEKIYCPPEASVLLASYAVQAKYGDYDPSVHKR 160

Qy   86 CWLALDELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAEVRE 128
```

```
Db  161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201

RESULT 11
US-08-628-145-5
; Sequence 5, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
;   APPLICANT: Seizinger, Bernd R.
;   APPLICANT: Kley, Nikolai A.
;   APPLICANT: Bianchi, Albert B.
;   TITLE OF INVENTION: No. 5872214el NF2 Isoforms
;   NUMBER OF SEQUENCES: 26
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Reed & Robins
;     STREET: 635 Bryant Street
;     CITY: Palo Alto
;     STATE: California
;     COUNTRY: U.S.A
;     ZIP: 94301
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/628,145
;     FILING DATE: 04-APR-1996
;     CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/179,738
;     FILING DATE: 10-JAN-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Robins, Roberta L.
;     REGISTRATION NUMBER: 33,208
;     REFERENCE/DOCKET NUMBER: 5998-0017
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (415) 617-8999
;     TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 591 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FRAGMENT TYPE: N-terminal
;   ORIGINAL SOURCE:
;     ORGANISM: murine
US-08-628-145-5

Query Match   9.6%; Score 68.5; DB 2; Length 591;
Best Local Similarity  25.7%; Pred. No. 4;
Matches  27; Conservative  22; Mismatches  43; Indels  13; Gaps  4;

Qy   33 YNEQDEEAAVLEVPQHSDSLLLHCRIIE----ADPQTSITLYSMLLQL--NFEMAAMRG 85
Db  101 YPENAEELVQEITQHLFFLQVKKQILDEKVYCPPEASVLLASYAVQAKYGDYDPSVHKR 160
Qy   86 CWLALDELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAEVRE 128
Db  161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201

RESULT 12
US-08-628-145-10
; Sequence 10, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
;   APPLICANT: Seizinger, Bernd R.
;   APPLICANT: Kley, Nikolai A.
;   APPLICANT: Bianchi, Albert B.
;   TITLE OF INVENTION: No. 5872214el NF2 Isoforms
;   NUMBER OF SEQUENCES: 26
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Reed & Robins
;     STREET: 635 Bryant Street
;     CITY: Palo Alto
;     STATE: California
;     COUNTRY: U.S.A
;     ZIP: 94301
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/628,145
;     FILING DATE: 04-APR-1996
;     CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/179,738
;     FILING DATE: 10-JAN-1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Robins, Roberta L.
;     REGISTRATION NUMBER: 33,208
;     REFERENCE/DOCKET NUMBER: 5998-0017
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (415) 617-8999
;     TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 10:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 591 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   FRAGMENT TYPE: N-terminal
;   ORIGINAL SOURCE:
;     ORGANISM: Homo sapien
US-08-628-145-10

Query Match   9.6%; Score 68.5; DB 2; Length 591;
Best Local Similarity  25.7%; Pred. No. 4;
Matches  27; Conservative  22; Mismatches  43; Indels  13; Gaps  4;

Qy   33 YNEQDEEAAVLEVPQHSDSLLLHCRIIE----ADPQTSITLYSMLLQL--NFEMAAMRG 85
Db  101 YPENAEELVQEITQHLFFLQVKKQILDEKIYCPPEASVLLASYAVQAKYGDYDPSVHKR 160
Qy   86 CWLALDELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAEVRE 128
Db  161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201

RESULT 13
US-08-171-718-16
; Sequence 16, Application US/08171718
; Patent No. 5707863
; GENERAL INFORMATION:
;   APPLICANT: Trofatter, James A.
;   APPLICANT: MacCollin, Mia M.
;   APPLICANT: Gusella, James F.
;   TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses
;   TITLE OF INVENTION: Thereof
;   NUMBER OF SEQUENCES: 120
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;     STREET: 1100 New York Avenue, N.W., Suite 600
;     CITY: Washington
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20005-3934
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
```

```
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/171,718
;       FILING DATE: 22-DEC-1993
;       CLASSIFICATION: 436
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/108,808
;       FILING DATE: 19-AUG-1993
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/022,034
;       FILING DATE: 25-FEB-1993
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/026,063
;       FILING DATE: 04-MAR-1993
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Brown, Anne
;       REGISTRATION NUMBER: 36,463
;       REFERENCE/DOCKET NUMBER: 0609.3850003
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (202) 371-2600
;       TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 16:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 595 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;       MOLECULE TYPE: protein
US-08-171-718-16

Query Match          9.6%;      Score 68.5;  DB 1;  Length 595;
Best Local Similarity   25.7%;  Pred. No. 4.1;
Matches  27;  Conservative  22;  Mismatches  43;  Indels  13;  Gaps  4;

Qy   33 YNEQDEEAAVLEVPQHSDSLLLHCRIIE---ADPQTSITLYSMLLQL--NFEMAAMRG 85
            :||  :| :|| :|| ::|::|:: |::  :::: :::  : :::: :
Db  101 YPENAEEELVQEITQHLFFLQVKKQILDEKIYCPPEASVLLASYAVQAKYGDYDPSVHKR 160

Qy   86 CWLALDELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAEVRE 128
           :| :||  :|   :::|||:|: :|:: :::| |:
Db  161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201

RESULT 14
US-08-478-087-16
; Sequence 16, Application US/08478087
; Patent No. 6077685
;   GENERAL INFORMATION:
;       APPLICANT: Trofatter, James A.
;       APPLICANT: MacCollin, Mia M.
;       APPLICANT: Gusella, James F.
;       TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses
;       TITLE OF INVENTION: Thereof
;       NUMBER OF SEQUENCES: 120
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;       STREET: 1100 New York Avenue, N.W., Suite 600
;       CITY: Washington
;       STATE: D.C.
;       COUNTRY: USA
;       ZIP: 2005-3934
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/478,087
;       FILING DATE: 07-JUN-1995
;       CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:

;       APPLICATION NUMBER: US 08/171,718
;       FILING DATE: 22-DEC-1993
;       APPLICATION NUMBER: US 08/108,808
;       FILING DATE: 19-AUG-1993
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/022,034
;       FILING DATE: 25-FEB-1993
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/026,063
;       FILING DATE: 04-MAR-1993
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Brown, Anne
;       REGISTRATION NUMBER: 36,463
;       REFERENCE/DOCKET NUMBER: 0609.3850003
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (202) 371-2600
;       TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 16:
;   SEQUENCE CHARACTERISTICS:
;       LENGTH: 595 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;       MOLECULE TYPE: protein
US-08-478-087-16

Query Match          9.6%;      Score 68.5;  DB 3;  Length 595;
Best Local Similarity   25.7%;  Pred. No. 4.1;
Matches  27;  Conservative  22;  Mismatches  43;  Indels  13;  Gaps  4;

Qy   33 YNEQDEEAAVLEVPQHSDSLLLHCRIIE---ADPQTSITLYSMLLQL--NFEMAAMRG 85
            :||  :| :|| :|| ::|::|:: |::  :::: :::  : :::: :
Db  101 YPENAEEELVQEITQHLFFLQVKKQILDEKIYCPPEASVLLASYAVQAKYGDYDPSVHKR 160

Qy   86 CWLALDELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAEVRE 128
           :| :||  :|   :::|||:|: :|:: :::| |:
Db  161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201

RESULT 15
US-08-179-738-2
; Sequence 2, Application US/08179738
; Patent No. 557846-2
;   GENERAL INFORMATION:
;       APPLICANT: Seizinger, Bernd R.
;       APPLICANT: Kley, Nikolai A.
;       APPLICANT: Bianchi, Albert B.
;       TITLE OF INVENTION: No. 557846-2el NF2 Isoforms
;       NUMBER OF SEQUENCES: 26
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Reed & Robins
;       STREET: 635 Bryant Street
;       CITY: Palo Alto
;       STATE: California
;       COUNTRY: U.S.A
;       ZIP: 94301
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/08/179,738
;       FILING DATE: 10-JAN-1994
;       CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Robins, Roberta L.
;       REGISTRATION NUMBER: 33,208
;       REFERENCE/DOCKET NUMBER: 5998-0017
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: (415) 617-8999
;       TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 2:
```

```
;   SEQUENCE CHARACTERISTICS:
;       LENGTH:  596 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;       MOLECULE TYPE:  protein
;       FRAGMENT TYPE:  N-terminal
;       ORIGINAL SOURCE:
;       ORGANISM:  murine
US-08-179-738-2
```

```
Query Match          9.6%;  Score 68.5;  DB 1;  Length 596;
Best Local Similarity   25.7%;  Pred. No. 4.1;
Matches   27;  Conservative  22;  Mismatches  43;  Indels  13;  Gaps   4;

Qy   33 YNEQDEEAAVLEVPQHSDSLLLHCRIIE----ADPQTSITLYSMLLQL--NFEMAAMRG 85
           | |:|| | |:|| | :| :||::||     ||:||::| | :|  :||:: :  .:
Db  101 YPENAEEELVQEITQHLFFLQVKKQILDEKVYCPPEASVLLASYAVQAKYGDYDPSVHKR 160

Qy   86 CWLALDELHNVRL--CFQQSLEHLDEASFSDIVSGFIEHAAEVRE 128
           : |  | || ||  |:   :| ::| |      | | ||
Db  161 GFLAQEELLPKRVINLYQMTPEMWEER---ITAWYAEHRGRARD 201
```

```
Search completed: June  5, 2001, 18:17:00
Job time: 186 sec
```

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:16:36 ; Search time 55.68 Seconds
                                     (without alignments)
                                     142.703 Million cell updates/sec

Title: US-09-596-784-4
Perfect score: 715
Sequence: 1 MTSSQQRVERFLQYFSAGCK.........IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5
Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   A_Geneseq_0401:*
           1:   /SIDS6/gcgdata/geneseq/geneseqp/AA1980.DAT:*
           2:   /SIDS6/gcgdata/geneseq/geneseqp/AA1981.DAT:*
           3:   /SIDS6/gcgdata/geneseq/geneseqp/AA1982.DAT:*
           4:   /SIDS6/gcgdata/geneseq/geneseqp/AA1983.DAT:*
           5:   /SIDS6/gcgdata/geneseq/geneseqp/AA1984.DAT:*
           6:   /SIDS6/gcgdata/geneseq/geneseqp/AA1985.DAT:*
           7:   /SIDS6/gcgdata/geneseq/geneseqp/AA1986.DAT:*
           8:   /SIDS6/gcgdata/geneseq/geneseqp/AA1987.DAT:*
           9:   /SIDS6/gcgdata/geneseq/geneseqp/AA1988.DAT:*
          10:   /SIDS6/gcgdata/geneseq/geneseqp/AA1989.DAT:*
          11:   /SIDS6/gcgdata/geneseq/geneseqp/AA1990.DAT:*
          12:   /SIDS6/gcgdata/geneseq/geneseqp/AA1991.DAT:*
          13:   /SIDS6/gcgdata/geneseq/geneseqp/AA1992.DAT:*
          14:   /SIDS6/gcgdata/geneseq/geneseqp/AA1993.DAT:*
          15:   /SIDS6/gcgdata/geneseq/geneseqp/AA1994.DAT:*
          16:   /SIDS6/gcgdata/geneseq/geneseqp/AA1995.DAT:*
          17:   /SIDS6/gcgdata/geneseq/geneseqp/AA1996.DAT:*
          18:   /SIDS6/gcgdata/geneseq/geneseqp/AA1997.DAT:*
          19:   /SIDS6/gcgdata/geneseq/geneseqp/AA1998.DAT:*
          20:   /SIDS6/gcgdata/geneseq/geneseqp/AA1999.DAT:*
          21:   /SIDS6/gcgdata/geneseq/geneseqp/AA2000.DAT:*
          22:   /SIDS6/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	715	100.0	139	20	W98012	Hypersensitive res
2	715	100.0	139	21	Y71096	Erwinia amylovora
3	715	100.0	139	21	Y84857	A hypersensitive r
4	80.5	11.3	368	20	W89180	Amino acid sequenc
5	80.5	11.3	766	21	B07739	A snake venom prot
6	80.5	11.3	787	21	B07740	A snake venom prot
7	80.5	11.3	820	21	B07741	A snake venom prot
8	76.5	10.7	1105	19	W44864	Human TPC2 telomer
9	76.5	10.7	1105	20	W73958	Human TPC2 protein
10	75	10.5	243	19	W38542	Streptococcus pneu
11	75	10.5	317	19	Y85961	S. pneumoniae deri
12	75	10.5	317	21	B13516	Streptococcus pneu
13	72.5	10.1	466	22	B63634	Human gastric canc
14	71.5	10.0	920	19	W82500	Human OGT protein.
15	70.5	9.9	2183	19	W48708	Measles virus Mora
16	70.5	9.9	2183	19	W48704	Measles virus 1977
17	69.5	9.7	2183	14	R39592	L protein of atten
18	69.5	9.7	2183	19	W48706	Measles virus Mont
19	69.5	9.7	2183	19	W48707	Measles virus Rube
20	69.5	9.7	2183	19	W48709	Measles virus Zagr
21	69.5	9.7	2183	19	W48710	Measles virus AIK-
22	69.5	9.7	2183	19	W48703	Measles virus Edmo
23	69.5	9.6	2183	19	W48705	Measles virus 1983
24	68.5	9.6	239	20	Y25380	HPV fusion protein
25	68.5	9.6	239	20	Y02636	CLYTA-E7-His prote
26	68.5	9.6	584	20	W94457	Mouse neurofibroma
27	68.5	9.6	590	20	W94458	Human neurofibroma
28	68.5	9.6	591	18	W09646	Mouse merlin prote
29	68.5	9.6	591	18	W09648	Human merlin prote
30	68.5	9.6	591	18	W94456	Mouse neurofibroma
31	68.5	9.6	595	15	R60398	Merlin protein enc
32	68.5	9.6	596	18	W09645	Mouse merlin prote
33	68.5	9.6	596	20	W94455	Mouse neurofibroma
34	68.5	9.6	596	20	W94459	Human neurofibroma
35	68	9.5	225	20	Y34662	Chlamydia pneumoni
36	68	9.5	652	17	R88124	Tobacco mosaic vir
37	68	9.5	1087	20	Y19935	B. burgdorferi ant
38	68	9.5	1119	20	Y19934	B. burgdorferi ant
39	68	9.5	1144	17	R88122	Tobacco mosaic vir
40	68	9.5	1144	17	R88123	Tobacco mosaic vir
41	67.5	9.4	404	21	B20940	Drosophila odorant
42	67.5	9.4	584	20	W09647	Mouse merlin prote
43	67.5	9.4	1527	20	Y43543	A human MPR-relate
44	67.5	9.4	1528	18	W33363	Human multidrug re
45	67	9.4	1194	21	B36533	Preliminary CLASP-

ALIGNMENTS

RESULT   1
W98012
ID   W98012 standard; Protein; 139 AA.
XX
AC   W98012;
XX
DT   21-JUN-1999 (first entry)
XX
DE   Hypersensitive response elicitor DspF.
XX
KW   Hypersensitive response elicitor; DspF; disease resistance;
KW   insect resistance; biological control; transgenic plant.
XX
OS   Erwinia amylovora.
XX
PN   WO9907206-A1.
XX
PD   18-FEB-1999.
XX
PF   24-JUL-1998; 98WO-US15426.
XX
PR   06-AUG-1997; 97US-0055105.
XX
PA   (CORR) CORNELL RES FOUND INC.
XX
PI   Beer SV, Bogdanove AJ, Kim JF, Wei Z;
XX
DR   WPI; 1999-180362/15.
DR   N-PSDB; X24811.
XX
PT   Nucleic acid encoding hypersensitive response-eliciting protein -
PT   used to improve growth of plants and impart resistance to disease
PT   and insects
XX

PS   Claim 18; Page 57-58; 75pp; English.
XX
CC   This polypeptide comprises the 16 kDa hypersensitive response
CC   elicitor protein DspF of Erwinia amylovora. The nucleotide
CC   sequence of the dsp region of E. amylovora strain Ea321 was
CC   determined using subclones of pCPP430. A two-gene operon was
CC   discovered comprising dspE (see X24810) and dspF (see X24811). The
CC   isolated dsp DNA molecules and encoded proteins can be used to
CC   impart disease resistance to plants, to enhance plant growth,
CC   and/or to control insects on plants. This is achieved by applying
CC   a hypersensitive response elicitor protein or polypeptide in a
CC   non-infectious form to plants or plant seeds, or by producing
CC   transgenic plants or plant seeds transformed with DNA encoding a
CC   hypersensitive response elicitor. Protection can be provided
CC   against a wide range of viruses, bacteria, fungi and insects, e.g.
CC   tobacco mosaic virus and tomato mosaic virus. Pseudomonas syringae,
CC   Xanthomonas campestris, Fusarium oxysporum, Phytophthora infestans,
CC   armyworm, diamondback moth, etc. The method avoids use of
CC   infectious agents or polluting chemicals. Claimed transgenic
CC   plants are selected from alfalfa, rice, wheat, barley, rye, cotton,
CC   sunflower, peanut, corn, potato, bean, pea, chicory, lettuce,
CC   endive, cabbage, brussel sprout, sweet potato, beet, parsnip,
CC   turnip, cauliflower, broccoli, turnip, radish, spinach, onion,
CC   garlic, eggplant, pepper, celery, carrot, squash, pumpkin,
CC   zucchini, cucumber, apple, pear, melon, citrus, strawberry, grape,
CC   raspberry, pineapple, soybean, tobacco, tomato, sorghum, sugarcane,
CC   Arabidopsis thaliana, Sainpaulia, petunia, pelargonium, poinsettia,
CC   chrysanthemum, carnation and zinnia.
XX
SQ   Sequence 139 AA;

Query Match   100.0%; Score 715; DB 20; Length 139;
Best Local Similarity   100.0%; Pred. No. 9e-79;
Matches 139; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 mtssqqrverflqyfsagcktpihlkdgvcalyneqdeeaavlevpqhsdslllhcriie 60

Qy   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 adpqtsitlysmllqlnfemaamrgcwlaldelhnvrlcfqqslehldeasfsdivsgfi 120

Qy 121 EHAAEVREYIAQLDESSAA 139
        |||||||||||||||||||
Db 121 ehaaevreyiaqldessaa 139

RESULT 2
Y71096
ID   Y71096 standard; Protein; 139 AA.
XX
AC   Y71096;
XX
DT   08-SEP-2000 (first entry)
XX
DE   Erwinia amylovora hypersensitive response elicitor encoded by dspF gene.
XX
KW   Hypersensitive response elicitor; enviromental stress resistance;
KW   plant; dspF gene.
XX
OS   Erwinia amylovora.
XX
PN   WO200028055-A2.
XX
PD   18-MAY-2000.
XX
PF   04-NOV-1999; 99WO-US26039.
XX
PR   05-NOV-1998; 98US-0107243.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.

XX
PI   Wei Z, Schading RL;
XX
DR   WPI; 2000-376566/32.
DR   N-PSDB; D00671.
XX
PT   Application of a hypersensitive response elicitor protein to plants to
PT   impart stress resistance -
XX
PS   Disclosure; Page 21; 84pp; English.
XX
CC   The patent discloses a method to impart stress resistance to plants by
CC   applying a hypersensitive response elicitor in a non-infectious form to
CC   a plant or seed. The present sequence is a hypersensitive response
CC   elicitor encoded by dspF gene from Erwinia amylovora.
CC   The protein is used to impart stress resistance to plants.
XX
SQ   Sequence 139 AA;

Query Match   100.0%;   Score 715; DB 21; Length 139;
Best Local Similarity   100.0%; Pred. No. 9e-79;
Matches 139; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIE 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 mtssqqrverflqyfsagcktpihlkdgvcalyneqdeeaavlevpqhsdslllhcriie 60

Qy   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 adpqtsitlysmllqlnfemaamrgcwlaldelhnvrlcfqqslehldeasfsdivsgfi 120

Qy 121 EHAAEVREYIAQLDESSAA 139
        |||||||||||||||||||
Db 121 ehaaevreyiaqldessaa 139

RESULT 3
Y84857
ID   Y84857 standard; Protein; 139 AA.
XX
AC   Y84857;
XX
DT   08-AUG-2000 (first entry)
XX
DE   A hypersensitive response elicitor protein.
XX
KW   Hypersensitive response; insect control; disease resistance;
KW   hypersensitive response elicitor; plant growth; vegetable; crop;
KW   ornamental plant; dspF gene.
XX
OS   Erwinia amylovora.
XX
PN   WO200020452-A2.
XX
PD   13-APR-2000.
XX
PF   05-OCT-1999; 99WO-US23181.
XX
PR   05-OCT-1998; 98US-0103050.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z, Fan H, Niggemeyer JL;
XX
DR   WPI; 2000-303745/26.
DR   N-PSDB; A14941.
XX
PT   Hypersensitive response elicitor polypeptides useful for imparting
PT   enhanced growth, disease resistance and insect resistance to plants,
PT   especially vegetables and ornamental flowers -
XX
PS   Disclosure; Page 22-23; 100pp; English.

```
XX
CC   The present sequence represents a hypersensitive response elicitor
CC   polypeptide. The polynucleotide represents the dspF gene. The
CC   specification describes hypersensitive response elicitor polypeptide
CC   fragments, which do not elicit a hypersensitive response. Instead,
CC   the proteins impart disease resistance to plants, enhance plant
CC   growth, and/or control insects. The gene can be used to
CC   used to these properties to plants. The polypeptide fragments may be
CC   in this way include vegetables, crops and ornamental plants such as
CC   alfalfa, rice, wheat, barley, rye, cotton, sunflower, peanut, corn,
CC   potato, sweet potato, bean, pea chicory, lettuce, endive, cabbage,
CC   brussel sprout, beet, parsnip, turnip, cauliflower, broccoli, radish,
CC   spinach, onion, garlic, eggplant, pepper, celery, carrot, squash,
CC   pumpkin, zucchini, cucumber, apple, pear, melon, citrus, strawberry,
CC   grape, raspberry, pineapple, soybean, tobacco, tomato, sorghum or
CC   sugarcane, Arabidopsis thaliana, Saintpaulia, petunia, pelargonium,
CC   poinsettia, chrysanthemum, carnation or zinnia.
XX
SQ   Sequence  139 AA;

Query Match             100.0%;  Score 715;  DB 21;  Length 139;
Best Local Similarity   100.0%;  Pred. No. 9e-79;
Matches 139;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAAVLEVPQHSDSLLHCRIIE 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 mtssqqrverflqyfsagcktpihlkdgvcalyneqdeeaavlevpqhsdsllhcriie 60

Qy   61 ADPQTSITLYSMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 adpqtsitlysmllqlnfemaamrgcwlaldelhnvrlcfqqslehldeasfsdivsgfi 120

Qy  121 EHAAEVREYIAQLDESSAA 139
        |||||||||||||||||||
Db  121 ehaaevreyiaqldessaa 139

RESULT  4
W89180
ID   W89180 standard; Protein; 368 AA.
XX
AC   W89180;
XX
DT   11-MAR-1999 (first entry)
XX
DE   Amino acid sequence of delta9 14:0-ACP desaturase.
XX
KW   delta 9 14:0-ACP desaturase; fatty acid; anacardic acid; pest;
KW   resistance; plant; pelargonium; geranium; polymer; nylon.
XX
OS   Pelargonium sp.
XX
PN   US5856157-A.
XX
PD   05-JAN-1999.
XX
PF   04-JUN-1997; 97US-0869137.
XX
PR   04-JUN-1996; 96US-0018957.
PR   04-JUN-1997; 97US-0869137.
XX
PA   (PENN-) PENN STATE RES FOUND.
XX
PI   Cox-foster DL, Craig R, Medford JI, Mumma RO, Schultz D;
XX
DR   WPI; 1999-105114/09.
DR   N-PSDB; V81284.
XX
PT   New isolated 14:0-ACP desaturase gene - obtained from geranium
PT   plants, used to enhance pest resistance of plants and for enhancing
PT   production of unsaturated fatty acids in plants
XX
PS   Claim 1; Fig 3; 26pp; English.
XX
CC   This represents a delta9 14:0-ACP desaturase. The 14:0-ACP desaturase
CC   gene enhances the accumulation 16:1 Delta-11 and 18:1 Delta-3 fatty acids
CC   and 22:1 omega 5 and 24:1 omega 5 anacardic acids. The gene can be used
CC   to provide pest resistance in plants. The gene can also be used to
CC   enhance the production of unsaturated fatty acids in plants such as
CC   soybeans, rapeseed, maize, sunflower, safflower, cotton, cuphea, peanut,
CC   coconut, oil-palm and pelargonium. The gene and its expression products
CC   can also be used in manufacturing methods, e.g. in innovative processes
CC   for making specialty polymers such as nylon and other methods in which
CC   unsaturated fatty acids are used as constituents or starting materials.
CC   In addition the promoter from the gene can be used for trichome specific
CC   expression of polypeptides.
XX
SQ   Sequence  368 AA;

Query Match             11.3%;  Score 80.5;  DB 20;  Length 368;
Best Local Similarity   20.1%;  Pred. No. 0.24;
Matches 37;  Conservative   26;  Mismatches   54;  Indels   67;  Gaps    8;

Qy    6 QRVERFLQYFSA-----GC-KTPIHLKDGVCALYNEQDEEAAVLE------VPQHSDSL 52
        ::|:| :||: |      ||  | || | ::|| :||     |  |        || |
Db  169 rqvektiqylialgqdigteknpyhl----fiytsfqeratfishantaklaqqhgdkq 223

Qy   53 LLH-CRIIEADPQTSITLYSMLLQLNFEM---------------- 80
        ||  |
Db  224 laqicgtiaadekrhetaytirvdklfeldpdetmsclahmmkrkitmpahlmrdgrdph 283

Qy   81 ------AAMRGCWLALDELHNVRLCFQQSLEH-LDEASFSDIVSGFIEHAAEVREYIA 131
          :  :: | | :  :|  ::|:       :::|
Db  284 lfqhfsvvasrtgvytvmdyin------ilehfvekwniekitaglsdkgreaqdyvc 335

Qy  132 QLDE 135
        :| |
Db  336 klge 339

RESULT  5
B07739
ID   B07739 standard; Protein; 766 AA.
XX
AC   B07739;
XX
DT   07-NOV-2000 (first entry)
XX
DE   A snake venom protease (SVPH) polypeptide varaint SVPH-1a.
XX
KW   Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin;
KW   chromosome 1; chromosome 4; immune system; splice variant.
XX
OS   Homo sapiens.
XX
PN   WO200043525-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-US01338.
XX
PR   21-JAN-1999;  99US-0116670.
PR   14-JUN-1999;  99US-0138682.
PR   27-SEP-1999;  99US-0155798.
XX
PA   (IMMV ) IMMUNEX CORP.
XX
PI   Cerretti DP;
XX
DR   WPI; 2000-482914/42.
DR   N-PSDB; A59304.
XX
PT   Snake venom protease (SVPH) nucleic acids, and polypeptides, used to
PT   identify proteins having metalloproteinase-disintegrin activity, and
```

PT   inhibitors of the proteins for use in therapeutics -
XX
PS   Claim 13; Page 15; 105pp; English.
XX
CC   The present sequence represnets an alternatively spliced snake venom
CC   protease-1 (SVPH-1) polypeptide, designated SVPH-1a. The SVPH
CC   polypeptides are metalloproteinase-disintegrin protein family members.
CC   The SVPH polynucleotides can be used as probes to identify nucleic acids
CC   encoding proteins having metalloproteinase-disintegrin activity, to
CC   identify human chromosome 1 or 4, to map genes on those chromosomes,
CC   to identify genes associated with diseases, syndromes and conditions
CC   associated with the chromosomes, and to study proteinases and their
CC   activities on cell/cell interactions and the immune system. Sense or
CC   antisense oligonucleotides of SVPH can be used to inhibit gene
CC   expression of SVPH 1, 3, or 4. The SVPH polypeptides can be used to
CC   study cell/cell and cell/matrix interactions involved in cellular
CC   processes and in the immune system. The polypeptides may also be used to
CC   screen for inhibitors of the polypeptide's activity, which are used in
CC   therapeutics. The antibodies can be used in assays to detect the
CC   presence of the polypeptides in vitro or in vivo, and to purify the
CC   polypeptides by affinity chromatography.
XX
SQ   Sequence   766 AA;

Query Match   11.3%;   Score 80.5; DB 21; Length 766;
Best Local Similarity   22.3%;   Pred. No. 0.67;
Matches   31; Conservative   28; Mismatches   47; Indels   33; Gaps   6;

Qy   18 GCKTPIHLK-----DGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITL-- 69
           | ||:|:   :   :| | : :| ||||||||: ::|
Db   60 gqkhiihikvkllfskhlpvftydqgailedqpfvqnncyyh-gyvegdpeslvslst 118

Qy   70 ----YSMLLQLNFEMAAMRGCWLALDELHNVR-LCFQQSLEHL------DEASFSDIVSG 118
              : ||| || | | || :| || ||
Db   119 cfgfqgilqin---------dfayeikplafstfehlvykmdseekqfstmrsg 165

Qy   119 FIEHAAEVREYIAQLDESS 137
           |::: | :: |:|:
Db   166 fmqneitcrmefeeidnst 184

RESULT   6
B07740
ID   B07740 standard; Protein; 787 AA.
XX
AC   B07740;
XX
DT   07-NOV-2000 (first entry)
XX
DE   A snake venom protease (SVPH-1) polypeptide varaint SVPH-1b.
XX
KW   Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin;
KW   chromosome 1; chromosome 4; immune system; splice variant.
XX
OS   Homo sapiens.
XX
PN   WO200043525-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-US01338.
XX
PR   21-JAN-1999; 99US-0116670.
PR   14-JUN-1999; 99US-0138682.
PR   27-SEP-1999; 99US-0155798.
XX
PA   (IMMV) IMMUNEX CORP.
XX
PI   Cerretti DP;
XX
DR   WPI; 2000-482914/42.
DR   N-PSDB: A59305.

XX
PT   Snake venom protease (SVPH) nucleic acids, and polypeptides, used to
PT   identify proteins having metalloproteinase-disintegrin activity, and
PT   inhibitors of the proteins for use in therapeutics -
XX
PS   Claim 13; Page 15-16; 105pp; English.
XX
CC   The present sequence represnets an alternatively spliced snake venom
CC   protease-1 (SVPH-1) polypeptide, designated SVPH-1b. The SVPH
CC   polypeptides are metalloproteinase-disintegrin protein family members.
CC   The SVPH polynucleotides can be used as probes to identify nucleic acids
CC   encoding proteins having metalloproteinase-disintegrin activity, to
CC   identify human chromosome 1 or 4, to map genes on those chromosomes,
CC   to identify genes associated with diseases, syndromes and conditions
CC   associated with the chromosomes, and to study proteinases and their
CC   activities on cell/cell interactions and the immune system. Sense or
CC   antisense oligonucleotides of SVPH can be used to inhibit gene
CC   expression of SVPH 1, 3, or 4. The SVPH polypeptides can be used to
CC   study cell/cell and cell/matrix interactions involved in cellular
CC   processes and in the immune system. The polypeptides may also be used to
CC   screen for inhibitors of the polypeptide's activity, which are used in
CC   therapeutics. The antibodies can be used in assays to detect the
CC   presence of the polypeptides in vitro or in vivo, and to purify the
CC   polypeptides by affinity chromatography.
XX
SQ   Sequence   787 AA;

Query Match   11.3%;   Score 80.5; DB 21; Length 787;
Best Local Similarity   22.3%;   Pred. No. 0.7;
Matches   31; Conservative   28; Mismatches   47; Indels   33; Gaps   6;

Qy   18 GCKTPIHLK-----DGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITL- 69
           | ||:|:   :   :| | : :| ||||||||: ::|
Db   60 gqkhiihikvkllfskhlpvftydqgailedqpfvqnncyyh-gyvegdpeslvslst 118

Qy   70 ----YSMLLQLNFEMAAMRGCWLALDELHNVR-LCFQQSLEHL------DEASFSDIVSG 118
              : ||| || | | || :| || ||
Db   119 cfgfqgilqin---------dfayeikplafstfehlvykmdseekqfstmrsg 165

Qy   119 FIEHAAEVREYIAQLDESS 137
           |::: | :: |:|:
Db   166 fmqneitcrmefeeidnst 184

RESULT   7
B07741
ID   B07741 standard; Protein; 820 AA.
XX
AC   B07741;
XX
DT   07-NOV-2000 (first entry)
XX
DE   A snake venom protease (SVPH-1) polypeptide varaint SVPH-1c.
XX
KW   Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin;
KW   chromosome 1; chromosome 4; immune system; splice variant.
XX
OS   Homo sapiens.
XX
PN   WO200043525-A2.
XX
PD   27-JUL-2000.
XX
PF   21-JAN-2000; 2000WO-US01338.
XX
PR   21-JAN-1999; 99US-0116670.
PR   14-JUN-1999; 99US-0138682.
PR   27-SEP-1999; 99US-0155798.
XX
PA   (IMMV) IMMUNEX CORP.
XX
PI   Cerretti DP;

```
XX
DR   WPI; 2000-482914/42.
DR   N-PSDB; A59306.
XX
PT   Snake venom protease (SVPH) nucleic acids, and polypeptides, used to
PT   identify proteins having metalloproteinase-disintegrin activity, and
PT   inhibitors of the proteins for use in therapeutics -
XX
PS   Claim 13; Page 16; 105pp; English.
XX
CC   The present sequence represnets an alternatively spliced snake venom
CC   protease-1 (SVPH-1) polypeptide, designated SVPH-1c. The SVPH
CC   polypeptides are metalloproteinase-disintegrin protein family members.
CC   The SVPH polynucleotides can be used as probes to identify nucleic acids
CC   encoding proteins having metalloproteinase-disintegrin activity, to
CC   identify human chromosome 1 or 4, to map genes on those chromosomes,
CC   to identify genes associated with diseases, syndromes and conditions
CC   associated with the chromosomes, and to study proteinases and their
CC   activities on cell/cell interactions and the immune system. Sense or
CC   antisense oligonucleotides of SVPH can be used to inhibit gene
CC   expression of SVPH 1, 3, or 4. The SVPH polypeptides can be used to
CC   study cell/cell and cell/matrix interactions involved in cellular
CC   processes and in the immune system. The polypeptides may also be used to
CC   screen for inhibitors of the polypeptide's activity, which are used in
CC   therapeutics. The antibodies can be used in assays to detect the
CC   presence of the polypeptides in vitro or in vivo, and to purify the
CC   polypeptides by affinity chromatography.
XX
SQ   Sequence  820 AA;

Query Match             11.3%;    Score 80.5;  DB 21;  Length 820;
Best Local Similarity   22.3%;    Pred. No. 0.74;
Matches  31;  Conservative  28;  Mismatches  47;  Indels  33;  Gaps  6;

Qy   18 GCKTPIHLK------DGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITL--  69
          :::|  ||:|     :| :| :::  |  | || :::|| : | |:: :||: :|
Db   60 gqkhiihikvxkllfskhlpvftytdggailedqpfvqnncyyh-gyvegdpeslvslst 118

Qy   70 ----YSMLQLNFEMAAMRGCWLALDELHNVR-LCFQQSLEHL------DEASFSDIVSG 118
             |||:                | :| |    | : :||    :|    : | ::||
Db  119 cfgfqgglqin---------dfayeikplafsttfehlvykmdseekqfstmrsg 165

Qy  119 FIEHAAEVREYIAQLDESS 137
          |:::  | :  ::|:
Db  166 fmqneitcrmefeeidnst 184


RESULT  8
W44864
ID   W44864 standard; Protein; 1105 AA.
XX
AC   W44864;
XX
DT   28-AUG-1998  (first entry)
XX
DE   Human TPC2 telomere length and telomerase regulatory protein.
XX
KW   TPC2; telomere length; telomerase; human; cancer; gene therapy;
KW   diagnosis; vaccine.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 182
FT                   /note= "encoded by TYT"
FT   Misc-difference 211
FT                   /note= "encoded by NGG"
FT   Misc-difference 217
FT                   /note= "encoded by GCT, C being uncertain"
FT   Misc-difference 231
FT                   /note= "encoded by GCT, C being uncertain"
FT   Misc-difference 381
FT                   /note= "encoded by ATC, C being uncertain"
FT   Misc-difference 383
FT                   /note= "encoded by TTA, T being uncertain"
FT   Misc-difference 387
FT                   /note= "encoded by AYT"
FT   Misc-difference 396
FT                   /note= "encoded by GGC, the second G being
FT                   uncertain"
FT   Misc-difference 396
FT                   /note= "encoded by GAG, the second G being
FT                   uncertain"
FT   Misc-difference 403
FT                   /note= "encoded by CYT"
FT   Misc-difference 404
FT                   /note= "encoded by TCA, C being uncertain"
FT   Misc-difference 409
FT                   /note= "encoded by GAA, G being uncertain"
FT   Misc-difference 410
FT                   /note= "encoded by CAG, G being uncertain"
FT   Misc-difference 413
FT                   /note= "encoded by ACT, C being uncertain"
FT   Misc-difference 416
FT                   /note= "encoded by TAC, A being uncertain"
FT   Misc-difference 419
FT                   /note= "encoded by GCY"
FT   Misc-difference 421
FT                   /note= "encoded by CYT"
FT   Misc-difference 426
FT                   /note= "encoded by CYT"
FT   Misc-difference 434
FT                   /note= "encoded by AGC, G being uncertain"
FT   Misc-difference 459
FT                   /note= "encoded by CTT, C being uncertain"
FT   Misc-difference 460
FT                   /note= "encoded by CCT, the second C being
FT                   uncertain"
FT   Misc-difference 464
FT                   /note= "encoded by CCT, the second C being
FT                   uncertain"
FT   Misc-difference 469
FT                   /note= "encoded by CCT, the second C being
FT                   uncertain"
FT   Misc-difference 476
FT                   /note= "encoded by AYT"
FT   Misc-difference 477
FT                   /note= "encoded by YTT"
FT   Misc-difference 487
FT                   /note= "encoded by GGY"
FT   Misc-difference 488
FT                   /note= "encoded by TCY"
FT   Misc-difference 491
FT                   /note= "encoded by CYT"
FT   Misc-difference 510
FT                   /note= "encoded by AAG, G being uncertain"
FT   Misc-difference 515
FT                   /note= "encoded by GAG, both Gs being uncertain"
FT   Misc-difference 517
FT                   /note= "encoded by CGN"
FT   Misc-difference 527
FT                   /note= "encoded by GYT"
FT   Misc-difference 529
FT                   /note= "encoded by CGA, G being uncertain"
FT   Misc-difference 530
FT                   /note= "encoded by GYG, the first G being
FT                   uncertain"
FT   Misc-difference 551
FT                   /note= "encoded by AGA, G being uncertain"
FT   Misc-difference 562
FT                   /note= "encoded by GCT, C being uncertain"
FT   Misc-difference 563
FT                   /note= "encoded by AGA, G being uncertain"
FT   Misc-difference 569
FT                   /note= "encoded by YTC, C being uncertain"
FT   Misc-difference 573
FT                   /note= "encoded by ATC, C being uncertain"
```

```
FT   Misc-difference 575
FT        /note= "encoded by YTC, C being uncertain"
FT   Misc-difference 578
FT        /note= "encoded by CCC, the third C being
FT             uncertain"
FT   Misc-difference 586
FT        /note= "encoded by YTT"
FT   Misc-difference 597
FT        /note= "encoded by GCG, the first G being
FT             uncertain"
FT   Misc-difference 599
FT        /note= "encoded by CGC, G being uncertain"
FT   Misc-difference 606
FT        /note= "encoded by GAR, G being uncertain"
FT   Misc-difference 608
FT        /note= "encoded by GAG, G being uncertain"
FT   Misc-difference 622
FT        /note= "encoded by AMC"
FT   Misc-difference 643
FT        /note= "encoded by AAG, G being uncertain"
FT   Misc-difference 645
FT        /note= "encoded by RAC, C being uncertain"
FT   Misc-difference 647
FT        /note= "encoded by GAG, G being uncertain"
XX
PN   WO9811204-A1.
XX
PD   19-MAR-1998.
XX
PF   13-SEP-1996;  96WO-US14679.
XX
PR   13-SEP-1996;  96WO-US14679.
XX
PA   (GERO-) GERON CORP.
XX
PI   Adams RR, Andrews WH, Feng J, Villeponteau B;
XX
DR   WPI; 1998-207373/18.
DR   N-PSDB; V19479.
XX
PT   Human TPC2, TPC3 and TR genes - regulate telomere length or modulate
PT   telomerase activity
XX
PS   Disclosure; Fig 4A-G; 86pp; English.
XX
CC   This polypeptide comprises human TPC2, a protein that regulates
CC   telomere length or modulates telomerase activity.  Its amino acid
CC   sequence was deduced from a cDNA clone (see V19479) obtained from a
CC   human 293 cell library.  TPC2 and TPC3 (see W44865) proteins can be
CC   obtained by purification from natural sources, by in vitro
CC   synthesis or by purification from recombinant host cells. They
CC   have application in methods for reconstituting in vitro telomerase
CC   or other enzymatic activities that maintain telomeres and regulate
CC   telomere length.  These methods have application in screens for
CC   therapeutic agents, and for diagnostic tests.  In addition,
CC   peptides corresponding to TPC2 or TPC3 proteins can also be used to
CC   regulate telomere length and telomerase activity in mammalian
CC   cells.  Immunogenic peptides and proteins of the invention can also
CC   be used in therapeutic immunisation and vaccination procedures.
CC   Antibodies that specifically bind to TPC2 or TPC3 proteins can
CC   be used in screening, diagnosing and monitoring diseases and other
CC   conditions, such as cancer, pregancy or fertility.
XX
SQ   Sequence  1105 AA;

Query Match             10.7%;  Score 76.5;  DB 19;  Length 1105;
Best Local Similarity   20.0%;  Pred. No. 3.4;
Matches   25;  Conservative   31;  Mismatches   46;  Indels   23;  Gaps  3;

Qy    25 LKDGVCALYNEQDEEAAVLEV-----------POHSDSLLHCRIIEADPQTSITLYSML 73
          |:|:|||  :|::|||
Db   691 ledkiralkenkdqlesvlevlhrqmeqyrdqpqhlekiayqqklliqed-----lvhir 744
```

```
Qy    74 LQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQL 133
          :|: |  |    |   :|:|: ::|::|||
Db   745 aelsrestemenawneylklendveqlkqtlgeqhrraf------ffqeksqiqkdlwri 798

Qy   134 DESSA 138
          :: :|
Db   799 edvta 803

RESULT   9
W73958
ID   W73958 standard; Protein; 1105 AA.
XX
AC   W73958;
XX
DT   29-APR-1999  (first entry)
XX
DE   Human TPC2 protein.
XX
KW   TPC2; TPC3; human; telomere length regulation; cancer; pregnancy;
KW   fertility; diagnosis; therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 1..1105
FT             /note= "X= unspecified amino acid"
XX
PN   US5858777-A.
XX
PD   12-JAN-1999.
XX
PF   13-SEP-1996;  96US-0710249.
XX
PR   08-SEP-1995;  95US-0003492.
PR   05-JAN-1996;  96US-0583808.
PR   13-SEP-1996;  96US-0710249.
XX
PA   (GERO-) GERON CORP.
XX
PI   Adams RR, Andrews WH, Feng J, Villeponteau B;
XX
DR   WPI; 1999-152104/13.
DR   N-PSDB; X01533.
XX
PT   DNA encoding proteins TPC2 and TPC3 - useful for regulating telomere
PT   length or modulating telomerase activity
XX
PS   Claim 1; Column 47-52; 59pp; English.
XX
CC   This sequence is the human TPC2 protein, which is contained within
CC   the recombinant mammalian host cell of the invention. The invention
CC   provides methods and reagents for regulating telomere length and
CC   modulating telomerase activity in mammalian cells as well as for
CC   detecting, diagnosing, and treating related diseases and conditions such
CC   as cancer, pregnancy, or fertility in humans and other mammals.
XX
SQ   Sequence  1105 AA;

Query Match             10.7%;  Score 76.5;  DB 20;  Length 1105;
Best Local Similarity   20.0%;  Pred. No. 3.4;
Matches   25;  Conservative   31;  Mismatches   46;  Indels   23;  Gaps  3;

Qy    25 LKDGVCALYNEQDEEAAVLEV-----------POHSDSLLHCRIIEADPQTSITLYSML 73
          |:|:|||  :|::|||
Db   691 ledkiralkenkdqlesvlevlhrqmeqyrdqpqhlekiayqqklliged-----lvhir 744

Qy    74 LQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFIEHAAEVREYIAQL 133
          :|: |  |    |   :|:|: ::|::|||
Db   745 aelsrestemenawneylklendveqlkqtlgeqhrraf------ffqeksqiqkdlwri 798
```

```
Qy    134 DESSA 138                                               Db    106 kkpissrdgmklcvetsttfddwvrqsekdyqdmliylkendfakigelte-knalamh- 163
          ::  :|                                                          |||    | ||   :||      .|| :::|||: :::||: .:||  | | .:|:::.:|
Db    799 edvta 803                                              Qy     57 RIIEADPQTSITLYSMLLQLNFE-MAAMR-------GCWLALDELHNVRL-CFQQSLEHL 107
                                                                          : :|:  .|| || :|| |||    | :|::|:      .:|: ||  .| | : :||| |||
RESULT   10                                                      Db    164 ---attktaspafsyltdasyeamafvrqlrekgeacyftmdagpnvkfcqekdlehl 219
W38542
ID   W38542 standard; Protein; 243 AA.                           Qy    108 DE 109
XX                                                                            |
AC   W38542;                                                     Db    220 se 221
XX
DT   06-NOV-1998 (first entry)                                   RESULT   11
XX                                                               Y85961
DE   Streptococcus pneumoniae protein of unknown function.       ID   Y85961 standard; Protein; 317 AA.
XX                                                               XX
KW   Streptococcus pneumoniae protein; genetic immunisation; antagonist;   AC   Y85961;
KW   immunological response; inoculation; antibody production; inhibitor;   XX
KW   T cell immune response; antimicrobial compound; bacterial adhesion;    DT   10-APR-2000 (first entry)
KW   extracellular matrix protein; protein-mediated cell invasion; wound;   XX
KW   pathogenesis.                                               DE   S. pneumoniae derived protein #170.
XX                                                               XX
OS   Streptococcus pneumoniae.                                   KW   Treatment; prevention; disease; diagnosis; gene therapy; screening;
XX                                                               KW   bacterial; antimicrobial; antibiotic; pathogenesis; infection.
PN   WO9743303-A1.                                               XX
XX                                                               OS   Streptococcus pneumoniae.
PD   20-NOV-1997.                                                XX
XX                                                               PN   WO9806734-A1.
PF   14-MAY-1997; 97WO-US07950.                                  XX
XX                                                               PD   19-FEB-1998.
PR   14-MAY-1996; 96US-0017670.                                  XX
XX                                                               PF   15-AUG-1997; 97WO-US14436.
PA   (SMIK ) SMITHKLINE BEECHAM CORP.                            XX
PA   (SMIK ) SMITHKLINE BEECHAM PLC.                             PR   16-AUG-1996; 96US-0024022.
XX                                                               XX
PI   Black MT, Hodgson JE, Knowles DJC, Nicholas RO;             PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PI   Stodola RK;                                                 XX
XX                                                               PI   Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
DR   WPI; 1998-008793/01.                                        PI   Stodola RK;
DR   N-PSDB; T98598.                                             XX
XX                                                               DR   WPI; 1998-159452/14.
PT   Novel Streptococcus pneumoniae proteins and related DNA - useful for   DR   N-PSDB; 296312.
PT   diagnosing anti-microbial agents for treatment of bacterial   XX
PT   infections                                                  PT   Streptococcus pneumoniae proteins and related DNA - useful for
XX                                                               PT   screening compounds for antibacterial activity
PS   Claim 12; Pages 320-321; 483pp; English.                    XX
XX                                                               PS   Claim 5; Page 456-457; 640pp; English.
CC   This sequence represents a Streptococcus pneumoniae protein of   XX
CC   unknown function, and is encoded by a DNA sequence of the invention.   CC   This invention describes novel isolated streptococcus pneumoniae
CC   The DNA sequences were isolated from Streptococcus pneumoniae strain   CC   polynucleotides (see 296173-296494) and their encoded proteins (see
CC   0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the   CC   Y85792-Y86182). The DNA, vectors and host cells described in the method
CC   invention can be used to identify compounds which interact with and   CC   of the invention are useful for the recombinant expression of the
CC   inhibit or activate the activity of the proteins. Antagonists can be   CC   polypeptides. The polypeptides are useful for treatment or prevention of
CC   used to treat diseases caused by S. pneumoniae proteins, through genetic   CC   disease, or diagnosis of disease related to expression or activity of
CC   immunisation. They can also be used to induce an immunological response   CC   such a polypeptide. They can also be used to screen for compounds which
CC   in a mammal by inoculation with the S. pneumoniae proteins or delivery   CC   interact with and inhibit or activate such a polypeptide. The
CC   of the encoding nucleic acids in a vector adequate to produce antibody   CC   polypeptides (or DNA encoding them, via gene therapy) are also useful
CC   and/or T cell immune responses to protect the animal from disease. The   CC   for inducing an immunological response in a mammal. The antagonists are
CC   proteins can also be used to identify antimicrobial compounds which are   CC   useful to inhibit such bacterial polypeptides. The polypeptides are
CC   capable of inhibiting their bioactivity. In particular the proteins of   CC   particularly useful to identify antimicrobial compounds and antibiotics.
CC   the invention can be used to prevent adhesion of bacteria to mammalian   CC   They are also useful to determine their role in pathogenesis of
CC   extracellular matrix proteins on in-dwelling devices or in wounds, to   CC   infection, dysfunction and disease.
CC   block protein-mediated mammalian cell invasion, and to block the normal   XX
CC   progression of pathogenesis in infections initiated other than by the   SQ   Sequence   317 AA;
CC   implantation of in-dwelling devices or other surgical techniques.
XX                                                               Query Match             10.5%;   Score 75;  DB 19;   Length 317;
SQ   Sequence   243 AA;                                          Best Local Similarity   23.8%;   Pred. No. 0.9;
                                                                 Matches   29;  Conservative   25;  Mismatches   30;  Indels   38;  Gaps   7;
Query Match             10.5%;   Score 75;  DB 19;   Length 243;
Best Local Similarity   23.8%;   Pred. No. 0.62;                 Qy     20 KTPIHLKDG--VCA-------------LYNEQDEEAAVLEVPQHSDSLLLHC 56
Matches   29;  Conservative   25;  Mismatches   30;  Indels   38;  Gaps   7;         |: || |:|   |:|          | | :|  :| :|
                                                                 Db    180 kkpissrdgmklcvetsttfddwvrqsekdyqdmliylkendfakigelte-knalamh- 237
Qy     20 KTPIHLKDG--VCA-------------LYNEQDEEAAVLEVPQHSDSLLLHC 56
```

```
Qy   57 RIIEADPQTSITLYSMLLQLNFE-MAAMR------GCWLALDELHNVRL-CFQQSLEHL 107
         ::|: :|| :|| | ||| :| ::||:| | ||::|  :||||
Db  238 ---attktaspafsyltdasyeamafvrqlrekgeacyftmdagpnvkvfcqekdlehl 293

Qy  108 DE 109
         |
Db  294 se 295

RESULT 12
B13516
ID   B13516 standard; Protein; 317 AA.
XX
AC   B13516;
XX
DT   02-NOV-2000 (first entry)
XX
DE   Streptococcus pneumoniae mevalonate diphosphate decarboxylase.
XX
KW   Mevalonate diphosphate decarboxylase; mvd; otitis media;
KW   conjunctivitis; pneumonia; bacteraemia; meningitis; sinusitis;
KW   pleural empyema; endocarditis; bacterial infection; Helicobacter pylori;
KW   stomach cancer; ulcer; gastritis.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200044764-A1.
XX
PD   03-AUG-2000.
XX
PF   19-JAN-2000; 2000WO-US01132.
XX
PR   28-JAN-1999; 99US-0238477.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Wilding EI, Gwynn M, Iordanescu M;
XX
DR   WPI; 2000-482967/42.
DR   N-PSDB; A63347.
XX
PT   New mvd (mevalonate diphosphate decarboxylase) polypeptide useful for
PT   identifying its agonists and antagonists which may be used for
PT   treatment and diagnosis of bacterial infections -
XX
PS   Claim 1; Page 3; 38pp; English.
XX
CC   The present sequence is the Streptococcus pneumoniae mevalonate
CC   diphosphate decarboxylase (mvd) protein. The protein, gene, agonists and
CC   antagonists can be used to diagnose and treat bacterial diseases,
CC   particularly those caused by Streptococcus pneumoniae, such as otitis
CC   media, conjunctivitis, pneumonia, bacteraemia, meningitis, sinusitis,
CC   pleural empyema and endocarditis, and those caused by Helicobacter
CC   pylori, including stomach cancer, ulcers and gastritis.
XX
SQ   Sequence 317 AA;

Query Match       10.5%;  Score 75;  DB 21;  Length 317;
Best Local Similarity 23.8%;  Pred. No. 0.9;
Matches 29;  Conservative 25;  Mismatches 30;  Indels 38;  Gaps 7;

Qy   20 KTPIHLKDG-VCA------LYNEQDEEAAVLEVPQHSDSLLLHC 56
          :| ::|: |        :| ::|  |||:  || ::||:
Db  180 kkpissrdgmklcvetstfddwvrqsekdydgdmliylkendfakigelte-knalamh- 237

Qy   57 RIIEADPQTSITLYSMLLQLNFE-MAAMR------GCWLALDELHNVRL-CFQQSLEHL 107
          ::|: :|| :|| | ||| :| ::||:| | ||::|  :||||
Db  238 ---attktaspafsyltdasyeamafvrqlrekgeacyftmdagpnvkvfcqekdlehl 293

Qy  108 DE 109
          |
```

```
Db  294 se 295

RESULT 13
B63634
ID   B63634 standard; Protein; 466 AA.
XX
AC   B63634;
XX
DT   26-MAR-2001 (first entry)
XX
DE   Human gastric cancer associated antigen protein sequence SEQ ID NO:996.
XX
KW   Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW   cancer associated antigen; cytostatic; cancer vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200073801-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14749.
XX
PR   28-MAY-1999;    99US-0136526.
PR   10-SEP-1999;    99US-0153454.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y;
XX
DR   WPI; 2001-025274/03.
XX
PT   Nucleic acids encoding breast, gastric and prostate cancer associated
PT   antigen precursors, useful for diagnosing and treating a condition
PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
PS   Example 1; Page 647-648; 799pp; English.
XX
CC   F22422 to F22626, F22627 to F22773 and F22774 to F23014 represent
CC   nucleotide sequences encoding human breast, gastric and prostate cancer
CC   associated antigen precursors (CAAP) respectively. B63232 to B63467,
CC   B63468 to B63721 and B63722 to B63970 represent human breast, gastric
CC   and prostate CAAP protein sequence respectively. CAAPs have cytostatic
CC   activity and can be used in the production of cancer vaccines. The human
CC   CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are
CC   useful for diagnosing and treating a condition characterised by
CC   expression of an abnormal amount of a protein, e.g. cancer.
XX
SQ   Sequence 466 AA;

Query Match       10.1%;  Score 72.5;  DB 22;  Length 466;
Best Local Similarity 26.0%;  Pred. No. 3.1;
Matches 32;  Conservative 21;  Mismatches 55;  Indels 15;  Gaps 5;

Qy   24 HLKDGVCALYNEQDEEAAVLEVPQHSDSLLLHCRIIEADPQTSITLYSMLLQLNFE---- 79
          :|: :|:|
Db  206 hekeaviclqnekdqklxemenimhsq----nceikelkgsreivled-lkklhvendek 260

Qy   80 MAAMRGCWLALDELHNVRLCFQQSLEHLDEASFSDIVS---GFIEHAAEVREYIAQLDE 135
          | | | :::
Db  261 lqliraelqsleqshlkeledtlqvrhiqe--fekvmdhrvsleelkkenqqiinqiqe 318

Qy  136 SSA 138
          |
Db  319 sha 321

RESULT 14
W82500
ID   W82500 standard; Protein; 920 AA.
```

```
XX
AC   W82500;
XX
DT   01-FEB-1999 (first entry)
XX
DE   Human OGT protein.
XX
KW   OGT; O-linked GlcNAc transferase; uridine; transferase; human; tumour;
KW   diphospho-N-acetylglucosamine; polypeptide beta-N-acetylglucosaminyl;
KW   predisposition; type II diabetes; hyperglycaemia; Alzheimer's disease;
KW   metastasis; diagnosis; inhibitor; treatment; diabetes mellitus.
XX
OS   Homo sapiens.
XX
PN   WO9844123-A2.
XX
PD   08-OCT-1998.
XX
PF   27-MAR-1998; 98WO-US06101.
XX
PR   31-MAR-1997; 97US-0042270.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Hanover JA, Lubas W;                        .
XX
DR   WPI; 1998-557118/47.
DR   N-PSDB; V69301.
XX
PT   Protein exhibiting O-linked GlcNAc transferase activity, OGT -
PT   useful, e.g. to assess predisposition to type II diabetes or
PT   Alzheimer's or metastatic potential of tumours, and to identify
PT   inhibitors
XX
PS   Claim 7; Page 31-33; 56pp; English.
XX
CC   This sequence represents a novel human O-linked GlcNAc transferase, OGT
CC   protein (also known as uridine diphospho-N-acetylglucosamine:
CC   polypeptide beta -N-acetylglucosaminyl transferase). This protein is
CC   useful to assess predisposition toward type II diabetes in patients
CC   suspected of having hyperglycaemia that could evolve into this disease,
CC   by assaying OGT activity in red blood cells. It can also be used to
CC   assess predisposition toward Alzheimer's disease, to assess the
CC   metastatic potential of tumours and to diagnose a tumour with metastatic
CC   potential. OGT can also be used to identify OGT inhibitors, especially in
CC   high-throughput assays, useful, e.g. in the treatment of diabetes
CC   mellitus, tumour-derived diseases and Alzheimer's disease.
XX
SQ   Sequence  920 AA;

Query Match        10.0%;  Score 71.5;  DB 19;  Length 920;
Best Local Similarity  24.4%;  Pred. No. 11;
Matches  30;  Conservative  21;  Mismatches  51;  Indels  21;  Gaps  4;

Qy   24  HLKDGVCALYNEQDEEAAVLEV-------------PQHSDSL--LLHCRIIEADPQTSITLY 70
             |  | | |  |||| :| :         |  |||   | | | :: :: :  || ||
Db  165  hfpdaycnlanalkekgsvaeaedcyntalrlcpthadslnnlanikreqgnieeavrly 224

Qy   71  SMLLQLNFEMAAMRGCWLALDELHNVRLCFQQSLEHLDEA----SFSDIVSGFIEHAAE 125
             ::  ||   |||| | ||| |  | : |||    |  |    ||  ||| ||| |||
Db  225  rkalevfpefaaahs--nlasvlqqqgklqealmhykeairisptfadaysnmgntlke 281

Qy  126  VRE 128
             :::
Db  282  mqd 284

RESULT 15
W48708
ID   W48708 standard; Protein; 2183 AA.
XX
AC   W48708;
XX
DT   13-OCT-1998 (first entry)
XX
DE   Measles virus Moraten vaccine L protein.
XX
KW   L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KW   single stranded RNA virus; Mononegavirales.
XX
OS   Measles virus.
XX
PN   WO9813501-A2.
XX
PD   02-APR-1998.
XX
PF   19-SEP-1997; 97WO-US16718.
XX
PR   27-SEP-1996; 96US-0026823.
XX
PA   (AMCY ) AMERICAN CYANAMID CO.
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI   Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
XX
DR   WPI; 1998-230710/20.
DR   N-PSDB; W48708.
XX
PT   Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT   single stranded RNA virus, of order Mononegavirales - having
PT   attenuating mutation in 3' genomic promoter region and RNA
PT   polymerase gene, useful as vaccine to immunise against such virus
XX
PS   Disclosure; Page 190-198; 426pp; English.
XX
CC   This sequence represents the L protein from Measles virus Moraten
CC   -vaccine. This sequence is used in a method which involves the
CC   isolation of recombinantly-generated, attenuated, non-segmented,
CC   negative-sense, single stranded RNA virus of the order Mononegavirales
CC   which have at least 1 attenuating mutation in the 3' genomic promoter
CC   region and at least 1 attenuating mutation in the RNA polymerase gene.
CC   This RNA virus can be used as a vaccine to immunise an individual against
CC   such a virus.
XX
SQ   Sequence  2183 AA;

Query Match        9.9%;  Score 70.5;  DB 19;  Length 2183;
Best Local Similarity  20.0%;  Pred. No. 48;
Matches  38;  Conservative  26;  Mismatches  63;  Indels  63;  Gaps  7;

Qy    4  SQQRVERFLQYFSAGCK----------------------TPIHLKDGVCALYNEQDEEAAVLEVP 46
             |  : | | :: :          ||:  |     |  ::
Db  158  ssqwfepflfwftvktemrsviksqthtchrrhtpvfftgssvellisrdlvaiiskes 217

Qy   47  QH----SDSLLLHCRIIEAD--PQTSITLYSMLLQLNFEMAAMRGCWLALD-------- 91
             :|     ||:|| :||| ||    :||| | :|||   |::|  ||  |
Db  218  qhvyyltfelvlmycdviegrlmtetamtidarytel---lgrvrymwklidgffpalgn 274

Qy   92  -----------------ELHNVRL-CFQQSLEHLDEASFSDIVSGFIEHAAE 125
                             ||| | ||  | | ||:|| || ||| ||| |||
Db  275  ptyqivamleplslaylqlrditvelrgaflnhcfteihdvldqngfsd--egtyhelte 332

Qy  126  VREYIAQLDE 135
             :||   :
Db  333  aldyifitdd 342

Search completed: June  5, 2001, 18:16:37
Job time: 203 sec
```

```
                    GenCore version 4.5
         Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:    June  5, 2001, 18:14:14 ; Search time 54.28 Seconds
                                      (without alignments)
                                      2327.059 Million cell updates/sec

Title:         US-09-596-784-2
Perfect score: 9448
Sequence:      1 MELKSLGTEHKAAVHTAAHN.........NPQVASALTDLKKEGLEMKS 1838
Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters:   198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     PIR_67:*
               1: pir1:*
               2: pir2:*
               3: pir3:*
               4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	9448	100.0	1838	2	T18448	pathogenicity fact
2	1824	19.3	1795	2	T30332	avirulence protein
3	275	2.9	2514	2	F81045	hemagglutinin/hemo
4	258.5	2.7	2248	2	A35938	profilaggrin - hum
5	255.5	2.7	3455	2	B82519	hemagglutinin-like
6	247	2.6	3442	2	E82589	hemagglutinin-like
7	245	2.6	2703	2	H81193	hemagglutinin/hemo
8	244.5	2.6	3259	1	A56539	giantin - human
9	243.5	2.6	3225	2	I52300	giantin - human
10	240.5	2.5	5327	2	T15564	microtubule-associ
11	237.5	2.5	4152	2	T31102	filamentous hemagg
12	232	2.5	1577	2	A35140	hemolysin A precur
13	229.5	2.4	2273	2	T09083	hemagglutinin/hemo
14	221	2.3	1608	2	A28182	hemolysin A - Serr
15	220.5	2.3	2253	2	T30336	nuclear/mitotic ap
16	219.5	2.3	2541	2	S11661	talin - mouse
17	218.5	2.3	4957	2	T03455	ALR protein - huma
18	218.5	2.3	5262	2	T03454	ALR protein - huma
19	217.5	2.3	1957	2	T38077	hypothetical coile
20	217	2.3	2712	2	T05113	hypothetical prote
21	214	2.3	3282	2	E82750	hemagglutinin-like
22	212.5	2.2	5627	2	C83339	hypothetical prote
23	210	2.2	2059	2	D82671	surface protein XF
24	209.5	2.2	1510	2	T33100	hypothetical prote
25	209.5	2.2	4919	2	T31105	hypothetical prote
26	207.5	2.2	3591	2	S21010	filamentous hemagg
27	205	2.2	1684	2	JW0057	gravin - human
28	204.5	2.2	4558	2	C82199	RTX toxin RtxA Vc1
29	204	2.2	2845	2	I49505	adenomatous polypo
30	203.5	2.2	1995	2	G81044	hemagglutinin/hemo
31	203.5	2.2	3187	2	JC5837	364K Golgi complex
32	202	2.1	1975	2	B81192	hemagglutinin/hemo
33	200	2.1	1467	2	A75564	conserved hypothet
34	199	2.1	2232	2	T34434	hypothetical prote
35	198.5	2.1	835	1	I57441	involucrin - orang
36	198	2.1	2015	2	B81989	hypothetical prote
37	197.5	2.1	1902	1	B44858	lactocepin (EC 3.4
38	195.5	2.1	4377	2	A55575	ankyrin 3, long sp
39	194	2.1	2954	2	T14156	kinesin-related pr
40	193.5	2.0	1962	2	A32634	lactocepin (EC 3.4
41	192.5	2.0	2938	2	T30249	cell proliferation
42	192	2.0	1217	2	S52714	sericiniB - silkwo
43	192	2.0	2346	2	T13829	Tpr homolog - frui
44	190.5	2.0	1536	2	A43855	high-molecular-wei
45	189	2.0	1902	2	S06997	lactocepin (EC 3.4

ALIGNMENTS

```
RESULT    1
T18448
pathogenicity factor DspA - Erwinia amylovora
C:Species: Erwinia amylovora
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T18448
R:Gaudriault, S.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z18936
A:Accession: T18448
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1838 <GAU>
A:Cross-references: EMBL:Y13831; PIDN:CAA74156.1
C:Genetics:
A:Note: dspA
C:Function:
A:Description: involved in pathogenicity

Query Match             100.0%;  Score 9448;  DB 2;  Length 1838;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1838;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy     1 MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQP 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MELKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQP 60

Qy    61 STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQH 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 STAADGISAAHQQKKSFSLRGCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQH 120

Qy   121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 EAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLS 180

Qy   181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 KMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDSEFQQ 240

Qy   241 LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 LHQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS 300

Qy   301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTOHYLAHHASSDGSQHLLLDNKG 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTOHYLAHHASSDGSQHLLLDNKG 360

Qy   361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQGTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420
```

```
Qy    421  PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKITIHLPELGVWQSADKDTHSQLSRQAD  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKITIHLPELGVWQSADKDTHSQLSRQAD  480

Qy    481  GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDAS  540

Qy    541  PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  PESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNEL  600

Qy    601  KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNFRQQHACPLGNDHQFHPGWNLT   660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  KMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNFRQQHACPLGNDHQFHPGWNLT   660

Qy    661  DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  DALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG  720

Qy    721  LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  LDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAM  780

Qy    781  AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  AVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALT  840

Qy    841  HEGEVFHQPREAWQNGAEESSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  HEGEVFHQPREAWQNGAEESSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG  900

Qy    901  GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKV  960

Qy    961  SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  SSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR 1020

Qy   1021  HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  HNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN 1080

Qy   1081  SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140

Qy   1141  KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  KGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQF 1200

Qy   1201  DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQG 1260

Qy   1261  SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVIPGAGITLDRAYN   1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  SAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVIPGAGITLDRAYN   1320

Qy   1321  LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  LSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDL 1380

Qy   1381  RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMQGSKLT  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  RIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMQGSKLT  1440

Qy   1441  FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500

Qy   1501  ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALDNRTSQSI   1560
```

```
Db   1501  ASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALDNRTSQSI   1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1561  SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAK 1620

Qy   1621  DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  DVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680

Qy   1681  AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  AALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVG 1740

Qy   1741  REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800

Qy   1801  QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838
           ||||||||||||||||||||||||||||||||||||||
Db   1801  QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS 1838


RESULT  2
T30332
avirulence protein - Pseudomonas syringae
C;Species: Pseudomonas syringae
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T30332
R;Bogdanove, A.J.; Kim, J.F.; Wei, Z.; Kolchinsky, P.; Charkowski, A.O.; Conlin, A.K.
Proc. Natl. Acad. Sci. U.S.A. 95, 1325-1330, 1998
A;Title: Homology and functional similarity of an hrp-linked pathogenicity locus, dsp
A;Reference number: 220825; MUID:98115919
A;Accession: T30332
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1795 <BOG>
A;Cross-references: EMBL:U97505; NID:g2978502; PIDN:AAC06134.1
C;Genetics:
A;Gene: avrE

Query Match            19.3%;  Score 1824;  DB 2;      Length 1795;
Best Local Similarity  29.1%;  Pred. No. 5.8e-83;
Matches  554;  Conservative 358;  Mismatches 722;  Indels 272;  Gaps 71;

Qy     34  SSSPQNAAASLAAEGKNRGKMPRIHQPSTAADGISAAHQQKKSFSLRGCLGTKKFSRSAP   93
            : :||  : |  :     |:: :  : ::|
Db     61  SKAPQQKAAT-PPTAKN-----VKTPPPASNVATPRNKARES---------GFSNVSSP  103

Qy     94  Q---------GQPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLTRSGGVKRRNMD  142
            |                 : ::                    |||        RKNLR  142
Db    104  QNTHRAPKWILRNHPNQASSGAQTHEI-------HPEAAP--------RKNLR       142

Qy    143  DMAGRPMVKGGSGGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIP 202
            :  :                                           DRLERSPSYLD  160
Db    143  VRFDLPQ----------                              DRLERSPSYLD   160

Qy    203  GSHHEIKEEPVGSTSKATTAHADRVEIAQEDDDEFQQLHQQRLARERENPPQP---PK   258
            ||:  || :::                    :  : | ||                 PK
Db    161  SDNPMTDEEAVANATRQF---------RSPDSHLQGSDGTRISMLATDPDPQPSSSGSK   209

Qy    259  LGVAT-PISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGSGAGVTPL-AVTLD-KGK  315
            :|  :  |     | : |   | ::  |   :  : |   : :   || :       KGK
Db    210  IGDSDGPIPPR----------EPMLWRSNGGRFELKDEKLVRNS----EPQGSIQLDAKGK 256

Qy    316  LQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVL  375
            : ||    : ::   |  :  :  ||:: ||| .: ::  || ||::|| :   :::   :
Db    257  PDFSTFNTPGLAPLLDSILATPKQTYLAHQSKDGVHGHQLLQANGHFLHLAQDDSSLAVI  316

Qy    376  HNSHPG-EIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQ  434
            :: :  | :|||:  |||  :  :  : :
```

```
Db   317 RSSNEALLIEGKKPPA-----VKMEREDGNIHI--DTASGRKTQ-ELPGKAHIAHITNV-  367

Qy   435 HPAGAARPQGESIRLHDDKIHILHPELGVWQSAD--KDT-HSQLSRQADGKLYALKDNRT  491
             :||    :.: ||| ||||: ||| |   |:  ||  ||| :|| |||| |
Db   368 ----LLSHDGERMRVHEDRLYQFDPISTRWKIPEGLEDTAFNSLSTGGNGSVYAKSDDAV  423

Qy   492 LQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHF  551
             : ||  .: ::|||:: | |   ||| |  | |   | |:  ||| |  |
Db   424 V-DLSSPFMPHVEVEDLQSFSVAPDNRAALLS----GKTTQAILLT-DMSP----VIGGLT  474

Qy   552 ADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALD  611
             :|| :|| ||| |||| || ::| | ||||||:  |:| |:||
Db   475 PKKTKGLELDGGKAQAAAVGLSGDKLFIADTQGRLYSAD-RSAFEGDDPKLKLMPEQANF  533

Qy   612 E---HFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLG-NDHQFHPGWNLTDALVID  666
             :        | ||  | || || ||.|:  |:| |||| :|:|| |   |||||
Db   534 QLEGVPLGGHNRVTGFINGDDGGVHALIKNRQGETHSHALDEQSSKLQSGWNLTNALVLN  593

Qy   667 NQLGLHHTNPEP---HEILDMGHLGSLALQEGKLHYFDQLTKGWTGAE-SDCKQLKKGLD  722
             | | |.  |      |::| |||| :|| :|:|||| |||  || |  |:|| :|:| ||
Db   594 NNRGL--TMPPPTAADRLNLDRAGLVGLSEGRIQRWDATPECWKDAGIKDIDRLQRGAD  651

Qy   723 GAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDKRA-QAMA  781
             || |||||| :||:    .:   || : |  || ::| |:|:| | :|||:|| :|| |
Db   652 SNAYVLKGGKLHALKIAAEHPNMAFDRNTALAQTARSTKVEMGKEIEGL--DDRVIKAFA  709

Qy   782 VIGVNKYLALTEKGDIRSFQIRPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALTH  841
             ||| ||| |   |:|| | || |::   :||||| | : :| |:|:|| || ||::
Db   710 MVSNKRFVALDD-------QNKLTAHSKDHKPVTLDIPGLEGDIKSLSLDEKHNLHALTS  762

Qy   842 EGEVFHQPREAWQN--GAE-SSSWHKLALPLQSESKLKSLDMSHEHKPIATFEDGSQH-  895
             |:|   | | | |   ||| | | || || :|:| ::| ||:|| |:||| |:|||||
Db   763 TGGLYCLPKEAWQSTKLGDQLRARWTPVALPGGQ-PVKALFTNDDNVLSAQIEDAEGKGL  821

Qy   896 -QLKAGGWHAYAAPERGQPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTG  954
             :|| ||| | |  |:| || | ||| |||:  |:|  | | :|| || | || | | |
Db   822 MQLKAGQWQRF--EQRP--VEENGLNDVHSRITGSNKTWRIPKTGLTILRMDVNTFGRSG  876

Qy   955 AEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHQWGREGLKPLYEMQGALIKQL  1014
             ||| || || |||| :||| | |:| |:|   ::: |:|| ||:|||   :| | |:|
Db   877 VEKSKKAST-SEFIRANIYKNTAETPRMKNVGDHIQHRYQGRIGLKEVYETESMLFKQL  935

Qy  1015 DAHNVRHNAPQP-------DLQSKLETLD--LGEHGAELLNDMKRFRDELEQSATRSVTVLG  1067
             |||: |  |||        | |:| ||    ||:   || |:|||| |:| :| :|: | |
Db   936 EL--IHESGGRPPARGQDLKARITALEAKLGPGGATLVKELETLRDELENHSYTALMSIG  993

Qy  1068 ------QHQGVLKSNGEINSEFKPSP-----GKALVQ---SFNVNRSGQDLSKSLQQ  1110
             :      ::||  :| ||| ||:||    :| | |     ||| ||| ||  || |||
Db   994 QSYGKAKNLKQQDGILNQHGEL---AKPSVRMQFGKKLADLGTKLNFKSSGHDLVKELQD  1050

Qy  1111 AVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLLLDTVTI  1170
             |: | : :| | |   ||| | |  : ||   :| ||:||: || || |||| || |: |
Db  1051 ALTQVAPSAENPTKKILLGTLKHQGLKLSHQKADIPLQQRRDASEDHGLSKARLAIDLVTL  1110

Qy  1171 GELHELADKAKLVSDHKPDADQIKQLRQQFDTFLREKRYESNPVKHYTDMGFTHNKALEAN  1230
             |:|  |:|:|   :: | | | |::| ||:| | : |:||:|||:|: |||| |||:| |
Db  1111 KSLGALLDQV----EQLPPQSDIEPLQKKLATILRDVTYGENPVKVVTDMGFTDNKALESG  1166

Qy  1231 YDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKIKNTLLSLDSG-ESMSFSRSYGG  1289
             |:| | ||| | : | |:|: :: ||| |::|:|||:|||  |::|   :|:| |:|| |
Db  1167 YESVKTFLKSFKKADHAVSVNMRAATGSKDQAELAGKFKSMLKQLEHGDDEVGLQRSYGV  1226

Qy  1290 GVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLN---VSFGRDGGVSGNIMV  1346
             | | : |||| ||  : ||  | ||: :|| ||:  :||||    || ::|| ||  |:|
Db  1227 NLTTPFIILADK--ATGLWPTAGATGNRNYIILNAERCEGGVTLYLISEGA-GNVSGGF--  1281

Qy  1347 ATGHDVMPYM-----TGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLFKK  1401
             :   |: ::|     | || ||:   | ||| |:  |   | || |  || | | |:| |
Db  1282 GAGKDYWPGFFDANNPARSVDVGN------NRTLTPNFRLGVDVTATVAASQRAGVVFN  1334

Qy  1402 LTEDELPGFIHGLTHGLTLPAELLQKGIEHQMKQGSKLTFSVDTSANLDLRAGINLNED-  1460
             ||| || || :|:|:||   | |:|  ::  |:|| ||   |: ||:| ||:|  |::|:
Db  1335 VPDEDIDAFVDDLFFEGQLNPLQVLKKAVDHESYEARRFNFDLTAGGTADIRAGINLTEDR  1394
```

```
Qy  1461 --GSKPN--GVTARVSAGLSA--SANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGA  1514
             :| ||   ||:|: || |:|   || |:| |::| | |||| ||:|:|||| || |||
Db  1395 DPNADPNSDFSAVVRGGFAANITVNLMTYDYSLTQKNDKTELKEGKNRPFLNNYTA  1454

Qy  1515 GANLTAALGVAHSSTHEGKPV--GIFPA--FTSTNVSAAL--ALDNRTSQSISLELKRA  1567
             |  || ||| | :  ||| ||  || |   | :: | :|   ||:|| | :| | |: |
Db  1455 GGQLRAQIGGSHTAP-TGTPASAGPTPASOTAANNLGGALNFSVENRTVKRIKFRYNVA  1513

Qy  1568 EPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDA----KPAE---QLHILQQHFS  1618
             :| |:| || | |:|| |:||:| ||: || |:|  |    ||:     :| | |:|
Db  1514 KPITTEGLSKLSKGLGEAFLDNTTKAKLAELADPLNARYTGKKPDEVIQAQLDGLEELFA  1573

Qy  1619 ----AKDVVGDERYEAVRNLKKLVIRQQA-ADSHSMELGSASHSTTYNNLSRINNDGIV-  1672
                 ||:| | :||||||:|:| ||::| ||:||:| || |||| :||| || |:|
Db  1574 DIPPPKD--NDKQYKALRDLKRAAVEHRASANKHSV-MDNARFETSKTNLSGLSSESLT  1630

Qy  1673 ELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEK  1732
             |||::||   | ||: || :| | |||:|||||||| ||  || ::| |:| || :|:|
Db  1631 KIMSSVRDASAP-GNATRVAEFMRQDPKLRAMLKEMEGSIGTLARVLEPKDSLVDKIDE  1689

Qy  1733 AILDGKVGREEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNI  1792
             |:|| : ||: |  |||||| |||| || | ||   | ||  ||| || || | |: |:|
Db  1690 GSLNGTMTQSDLSSMLEDRNEMRIKRLVVFHTATQAENFTSPTPLVSYNSGANVSVTKTL  1749

Qy  1793 GTINFKYGQDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMKS  1838
             | ||| |:| | |  |  ||| ::| || ||| |:: |||||||||
Db  1750 GRINFVYGADQDKPIGYTFDGELSRPSASLKEAAGDLKKEGFELKS  1795
```

RESULT 3
F81045

hemagglutinin/hemolysin-related protein NMB1768 [imported] - Neisseria meningitidis
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Dec-2000
C;Accession: F81045
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755
A;Accession: F81045
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2514 <TET>
A;Cross-references: GB:AE002526; GB:AE002098; NID:g7227015; PIDN:AAF42109.1; PID:g72.
A;Experimental source: serogroup B, strain MMD58
C;Genetics:
A;Gene: NMB1768

```
Query Match                     2.9%;  Score 275;  Pred. No. 3e-05;
Best Local Similarity  18.8%;  Conservative 287;  Mismatches 287;  Indels 586;  Gaps   91;
Matches 389;
```

```
Qy     7 GTEH-KAAVHTAAHNPV--------------GHGVALQQGSSS--------SSPQNAAAS  43
           |: ||  ||::|  | |               |:|| ::| |:|:        | | :|
Db    37 GSAHVKSVPFGTTHAPVCRSNIFSFSLLGFSLCLAVGTANIAFADGIIADKAAPKTQQAT  96

Qy    44 LAAEGKNRGKMPRIHQPSTAADGISAAHQQKFSLRGCLGTKKFSRSAPQGQPG------  98
             | :|||  || | || ||::|  || | ::|  || ||:| |  ||::| |
Db    97 ILQTGNG---IPQVNIQTPSAGVSVNQYAQFDVGNRGAILNN--SRSNTQTQLGGWIQG  151

Qy    99 ----------------------TTHSK----------GATLRDLLARDDGETQHEAAAPDAARLTR  132
                                   .::||          | | ||| | || || | | || | |
Db   152 NPWLARGEARVVNQINSSHSSOMNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATL  211

Qy   133 SGGVKRRNMDDMAGRPMVKG---------GSGED---------KVPTQQKRHQLNNFGQMRQTMLS  180
             ||| ||  | || | || |         |: |:          ||  ||  ||:|| |:|:|::
Db   212 TTGQPQYQAGDLSGFKIRQGNVVIAGHGLDARDTDFTRILSYHSKIDAPVWGQDVRVVAG  271
```

```
Qy   181  KMAHPASANAGDRLQHSPPHIPGSSHHEIKEEPVGSTSKATTAHADRVEIAQEDDSEFQQ   240
Db   272  QNDVVATGNA-----HSP------------ILNNAAANTSNNTANNGTHIPLFAIDTG-----  312

Qy   241  LHQQRLARERENPPQQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGS   300
Db   313  ------------KLG--GMYANKITLISTA------EQAGIRNQGQLFAS   342

Qy   301  GAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHAS--SDGSQHLLLDN   358
Db   343  SGN------VAIDANGRLVNSGTMAAAN------AKDTDNTAEHKVNIRSQG------VEN   385

Qy   359  KG-------HLFDIKSTATSYS----VLHNSHPGEIKGKLAQAGTGSVSVDGKSGKI   404
Db   386  SGTAVSQQGTQIHSQSIQNTGTLLSSGEILIHNS------GSLKNETSGTI----EAARL   435

Qy   405  SLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDKIHILHPELGVW   464
Db   436  AIDTDT-LNNQGKLSQTG------------SQKLHI   458

Qy   465  QSADKDTHSQLSRQADGKLYALKDNRTLQNLSD-----NKSSEKLVDKIKSYSVDQRGQ   518
Db   459  --------DAQGKM---DNRGRMGLQDTAPTASNGSSNQTGN--SYNASFHSS   498

Qy   519  VAILTDPPGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLV   578
Db   499  TTTPTTATGTGTATVSISNITAPT------FAD--GTIRTHGALDNSGSIIANGQTD   547

Qy   579  VADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN-   637
Db   548  VSAQQG-LNNAG---QIDIHQLNAKG---SAFDNHNG--TIISDAVHIQAGSLNNQNGNI   598

Qy   638  NFRQQHACPLGNDHQFHPGWNLTDALVID-NQLGLHHTNPEPHEILDMGHLGSLALQEGK   696
Db   599  TTRQQ----------------LEIETDQILDNAHGKILSAEIADLAVSGSLNNQNGE   638

Qy   697  LHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLP   756
Db   639  IATNQQL-----IIHDGQQSTAVIDNTNGTIQSGRDVAIQAK   675

Qy   757  HVRNKPEPGDALQGLNK----DDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLE   810
Db   676  SLSN----NGTLAADNKLDIALQDDFYVERNIVAGNE-LSLSTRGSLKNSHTLQAGKRIR   730

Qy   811  RPAQTLSREG----ISGELKDIIVDHKQNLYALTHEGEVFHQPREAWQNGAESSSWHKLA   866
Db   731  IKANNLDNAAQGNIQSGGTTDIGTQHN---LTNRGLIDGQ-------   767

Qy   867  LPQSESKLKSLDMSH---------EHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTS   918
Db   768  ----QTKIQAGQMNNIGTGRIYGDNIAIAATRLDNQDENGTGA--AIAARENLNLGIG--   819

Qy   919  GSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMS   978
Db   820  ----QLNNRENSLIYSGNDMAVGGALDTNGQATGKAQR------------   853

Qy   979  TPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQL--------DAHNVRHNAP   1024
Db   854  ----IHNAGATIEAAGKMRIGVERLHNTNEHLKTQLVETGREHIVDYEAFGRHELLREGT   909

Qy  1025  QPDL--------------QSKLETLDLGEH-------------GAELLNDMK   1049
Db   910  QHELGWSVYNDESDHLRTPDGAAHENWHKYDYEKVTQKTQVTQTAPAKIISGNDLLTIDGK   969

Qy  1050  R-FRDELEQSATRSVTVLGQHQG-----------VLKSNGEINSEFKPS-PGKALVQSFN   1096
Db   970  EVFNTDSQIIAGGNLIVQTEKDGLHNEQTFGEKKVFSENGKLHSYWREKHKGR---DSTG   1026

Qy  1097  VNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHEVSAGVDMSHQKGEIPLGRQRDPNDKT   1156
Db  1027  HSEQNYTLPEEITRNISLGSFAYESHRKALSHHAPSQGTELPQSNG---ISLPYTSNSFT   1083

Qy  1157  ALTKSRL-ILDTVTIGELHELADKAK----LVSDHKPDA--DQIKQLRQQFDTLREKR   1207
Db  1084  PLPSSSLYIINPVNKGYLVETDPRFANYRQWLGSDYMLDSLKLDPNNLHKRLGDGYYEQR   1143

Qy  1208  YESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKK   1267
Db  1144  LINEQIAELT-GHRRLDGYQNDEEQFKALMDNGATAARSMNLSVGIAL---SAEQVAQ   1197

Qy  1268  LKNTLLSLDSGESMSFSRSYGGGVSTVPVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTS   1327
Db  1198  LTSDIVWLVQKE----VKLPDGGTQTVLVPQVYVRVKNGDIDGKGAL-----LSGSNTQ   1247

Qy  1328  GGLNVSFGRDGGVSGNIMVATGHDVMPYMTGK---KTSA-----GNASDWLSAKHKIS   1377
Db  1248  INVSGGSLKNSGTIAGRNALIINTDTLDNIGGRIHAQKSAVTATQDINNIGGMLSAEQTLL   1307

Qy  1378  PDLRIGAAV-SGTLQGTLQNSLKFKLTEDELPG-FIHGLTHGTLT-----PAELLQKGIE   1430
Db  1308  --LNAGNNINSQSTTASSQNTQGSSTYLDRMAGLYIITGKEKGVLAAQAGKDINIIAGQIS   1365

Qy  1431  HQMKQGSKLTFSVDTSANLDL----RAGINLNEDGSKPNGVTARVSAGLSASA---   1479
Db  1366  NQSEQG-QTRLQAGRDINLDTVQTSKHQATHFDADNHVIRGSTNEVGSSIQTKGDVTLLS   1424

Qy  1480  --NLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVGI   1537
Db  1425  GNNLNAKAAEVSSANGTL--AVSAKND----IN-ISAGINTT--HVDDASKHTGRSGG-   1473

Qy  1538  FPAFTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTL---------GKHFK   1587
Db  1474  ---GNKLVITDKAQSHHETAQSSTFEGKQVVLQAGNDANLLGSNVISDNGTQIQAGNHVR   1530

Db  1588  DSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSH   1647

Db  1531  IGTT--------QTQSQSETHQTQKSGLMSAGIG-----FTI--GSKTN   1565

Qy  1648  SMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMM---NNDPALK   1702
Db  1566  TQENQSQSNEHTGSTVGSLKGDTTI-VAGKHYE------QIGSTVSSPEGNNTIYAQ   1615

Qy  1703  DIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVGREVGV-----LFQDRNNLRV   1756
Db  1616  SIDIQAAHNKLNSNTTQTYEQKGLTVAFSSPVTD-LAQQAIAVAQSSKQVGQSKND-RV   1672

Qy  1757  KSYSVSQS--VSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYGQDQNTPRRFT-LEG   1813
Db  1673  NAMAAANAGWQAYQTGKSAQNLANGTTNAKQVS------ISITYGEQQN-RQTTQVQA   1723

Qy  1814  GIAQANPQVASALTDLKKEGLEMKS   1838

Db  1724  NQAQASQIQAGGKTTLIATGAAEQS   1748
```

```
F;246-569/Region: filaggrin repeat
F;570-893/Region: filaggrin repeat
F;1074-1397/Region: filaggrin repeat
F;1573-1896/Region: filaggrin repeat


Query Match         2.7%; Score 258.5; DB 2; Length 2248;
Best Local Similarity  18.4%; Pred. No. 0.00017;
Matches  381; Conservative  249; Mismatches  815; Indels  629; Gaps  83;

Qy     7 GTEHKAAVHTAAHNPVGHGVALQ----------QGSSSSSPQNAAASLAAEGKNRGKMPR 56
           :    ::  :|:||          :|::::           || : |  | : |:
Db   487 GSGHQQSADSSRHSGIGHGQASSAVRDSGHRGYSGSGQASDQEGHSENSDSQSVSGQRQAR 546

Qy    57 IHQPSTAADGISAAHQQKKSFSLRGCLGTKKFSRSRSAPQGQPGTTHSKGATLRDLLARDDG 116
            ::|          :||                  ::||     | :|       |:::
Db   547 SHQQS----------HQES--------TRGQSGRSGRS---GSFLYQVSTHEQS 580

Qy   117 ETQHEAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGQMRQ 176
            :  ::: ||:|||   |: :| :||    :| |: ::|     ||   :|:|  ||:|:
Db   581 ESAHGRSAPSTRR-------------RQGSHHDQA-RDSSRHSASQEGQ-D 616

Qy   177 TMLSKMAHPASANAG-------DRLQHSPPHIPGSHH-----EIKEEPVGSTSKATT 221
            |:  : ||||||  :       || ||     ||:||      :|||||||||:|:|
Db   617 TI---RGHPGSSRGRQGSHYEQSVDRSGHS-----GSHHSHTTSQGRSDASHGTSGSRS 668

Qy   222 AHADRVEIAQEDDSEFQQLHQQRLARERENPPQPPKLQGVATPISARFQPKLTAVAESVL 281
            |  ||      :||||:||||:||||||:| |  |:|  ||  |:||||||| |||:|:
Db   669 ASRQTRDEQSGDGTRHSGSHHQEASTQADS-SRHSQVGQGQSAGSR-------TSRN 718

Qy   282 EGTDTTQ--SPLKPQSMLKGSGAGVT----------PLAVTLDKGKLQLAPDN 322
            :|:| |:   | |||| |||| ||:            ||| | |||| |||:|
Db   719 QGSSVSQDRDSEGQSEDSERHSGSASRNHRGSAQEQSRDGSRHPGSHDEDRAGHRQSADS 778

Qy   323 PPALNTLLKQTLGK----------DTQHYLAHHASSDGSQHLLLDNKGHLFDIKST 368
            || |:||:|||||:           || | |:||||:||:||||||:||||||||
Db   779 SSQSGTRHTQTSSRRQAASSQEQARSRAGDRHGSGHQQSADSSRH----------SGIGRG 829

Qy   369 ATSYSVLHNSHPGEIKGKLAQAGTGSVSVDKGSGKIS---LGSGTQSHNKTMLSQPGEA 424
            | |||:|::||| :| |||| || |||||| |||||     |||||||:|| |||||
Db   830 QASSAVRDRGHRGS-RGSGASDQEGH-SEDSDSQSVSAQOAGSHQQSHQEST---------- 880

Qy   425 HRSLLTGTWQHPAGAARPQGES----------IRLHDDKIHILHPELGVWQSADKDTHSQL 475
            ||| |||| ||| |||||||:            |:||::|||| |:| |||||||||||
Db   881 -----------RGRSQGRSGRSGSFIYQVSTHEQS----------ESAHGRTRTST 915

Qy   476 SRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMP 535
            |||| |||||||||||:||:|||:|||||| |||| || |:|:|||||||| |||| :|
Db   916 GRRQGSHHEQARDS--------SRHSASQEQDTIRAHPGSRRG----------GRQGSHHEQ 960

Qy   536 SLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQG 595
            |||||||||:|||||||||||||||||||||||||| |||| ||||||||||||||||
Db   961 GHSGSHHSHTTSQGRSDA-SHG----------QSG 989

Qy   596 DGNELKMKAMPQHALDEHFGHDHQISGFFHD-DHGQLNALVKNNFRQQHACPLGNDHQF 653
            :|:|:|:||||:||||||||||||||||||   |||||||:|||||||||  |||:|:
Db   990 SGSH----------HQQSADSSRHSGIGHGQASSAVRDSGHRGSSGSGQASDSEG 1033

Qy   654 HPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHY----FDQ 702
            ||| ||||||:|||||||||||||||||||||||||||||||||       |:
Db  1034 HSEDSDTQSVSAQOGQAGPHQ----QSHQESTRGRSAGRSGSFLYQVSTHEQSESAHGR 1090

Qy   703 ---LIFKGWTGA----------ESDCKQLKKGLDGAAYLLKDGEVKRLNIN----739
              |::|||||            |||||||||||||||||||||||||||:|
Db  1091 ARTSTRGROQGSHHEQARDSSRHSTSQEGQDTIRGHPGPSSGGRHGSHYEQSVNSTGHSGS 1150

Qy   740 ----QSTSSIKHGTENVFSLP-HVRNKPEPGDALQGLNK--------DDKAQAMAV 782
             :| ||||||||||||||||  ||||||||||||||||         |||||||:|
Db  1151 HHSHTTSQGRSDASHGTSGSRSASRETHNEEQSGDGSRHSGSRHQEASSWADSSGHSQAG 1210

Qy   783 IGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIH--VDHKQNLYALT 840
            :| :|:|||| ||||| |||||||||||||||||||||||||||||   |||||:|||||
Db  1211 QGESSG-SRTSRNQGSSF-----SQDSDSQGQSEDSERRSGSASRNHRGSAREQSRDGSR 1264

Qy   841 HEGEVFHQPREAWQNGAESSSWWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKA- 899
            ||| ::|| |: :|:|| ||:  ||:|||:|: ||:|::  |: ||  | ::||:|::||
Db  1265 HPGS-HHEDRAGHGHSADSSRQSGTRHTQTSSRRQAASSQEQARSRAGDRHGSGHQQSAD 1323

Qy   900 GGWHAYAAPERGPLAV-----GTSGSQTVFNR----LMQGVKGKVIPGS---GL 941
            ||| |||||||||| :     |||||||||||     |:||| |||||||   |
Db  1324 SSRHSGIGRGQASTAVRDSGHRGSRGSQASDNEGHSEDSDTQSVAGQRQAGSHHESHQES 1383

Qy   942 TVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLK 1001
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1384 TRGQSRETSGRSGSFLYQVST----------HEQSESSHGWSGTS--- 1418

Qy  1002 PLYEMQGALIKQLDAHNVRHNAPQPDLQSKLETLDL---------GEHGAELLNDMKFRFDELE 1056
            || ||||||||||||| ||||:|||||||||||||          |||||||||||||||||
Db  1419 -TRGRQGSHHEQAQ-DSSRHSASQ----------EGQDTIHGPGGHSA----------1459

Qy  1057 QSATRSVTVLGQHGVLKSNGEINSEFKPSPGKALVQSFNVNRS---GQDLSKSL--- 1108
            |||||||||||||||||||||||||||||||||||||||||||    |||:|||
Db  1460 QSGTRHTESSSRGQAA-SSHEQARSSAGERHGSHHQQSADSSRHAGIGHGQASSAVRDSG-D 1518

Qy  1109 ----------QQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSR 1162
                      |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1519 HRGYRGSQATDSEGHSEDSDTQSVSAQOQAGPHQQAHQ-ESARGQSGESSG----RSG 1571

Qy  1163 LILDTVTIGELHELADKAKLVS----------DHKPDADQIKQLRQQFDTLREKRYESNPV 1213
            ||||||||||||||||||||:           |||||||||||||||||||||||||||
Db  1572 SFLYQVSTHEQSESTHGQSVPSTGGRQGSHHDQAQDSSRHSASQEGQDTIRGHPGSSR-- 1629

Qy  1214 KHYTDMGFTHNKALEANYDAVKAFINAFKKEEHGVNLTTRTVLESQGSAELAKKLKNTLL 1273
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1630 -----GGRHGSHYEQSVDR----------SGHSGSHHSHTTSQGRSDASHGSGSRSASRQTRN 1678

Qy  1274 SLDSGESMSFSRSY----------GGVSTVFVPTLSKKVPVPVIPGAGITLD- 1316
            ||| ||||||||||          ||||||||||||||||||||||||||||
Db  1679 EEQSGDGSRHSRSHHHEASTQAESSRHSQAQGQSA-GPRTSRN----------QGSSVSQDS 1730

Qy  1317 ----------RAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNAS 1367
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1731 DSQGHSEDSERWSGGASRNHRGSAQEQSRDGS-----RHPTSHH-EDRAGHRHSA-ESS 1782

Qy  1368 DWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPGFIH----------GLTHG 1417
            ||||||||||||||||||||||||||||||||||||||||||||          |||||
Db  1783 DQSGTHH-AENSSGGOAASSHEQA----------RSSAGERHGSHHQQSADSSRHSGIGHG 1832

Qy  1418 TLTPAELLQKGIEHOMKQGSKLTFSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSA 1477
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1833 QASSA---VRDSGHRGSSGSQOASDSEGHSEDSDTQ----------SV 1866

Qy  1478 SANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVA----------1525
            |||||||||||||||||||||||||||||||||||||||||||||||
Db  1867 SAHQOAGSHQQSHQESTRGRSQEGSRGSFLYQVSTHEQESAHGRAGPSTGGRQGSRH 1926

Qy  1526 --------HSSTHEGKPV---------GIFPAFTSTNVSAAL 1550
                    |||||||||          |||||||||||||||
Db  1927 EQARDSSRHSASQEGQDTIRGHPGSRRGRQGSYHEQSVDRSGHSGSHHSHTTSQGRSDA 1986

Qy  1551 ALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQL 1610
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1987 SHQGSGSRSASRETRNEE-QSGDGSRHS--GSRHHEASTQADSSRHSQSGQGESAGSR 2041

Qy  1611 HILQQHFSAKDVVGDERY----------EAVRNLKKLVIRQQAADSHSMELGSASHSTTYNN 1662
            |||||||||||||||||           ||||||||||||||||||||||||||||||||
Db  2042 RSRROGSVSQDSDSEAYPEDSERRSESASRNHHGSSREQSRDG-SRHPGS-SHRDTASH 2099

Qy  1663 L--SRINNDGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSM 1720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2100 VQSSPVQSDSSTAKEHGHF-SSLSQDSAYHSG----------IQSRGSPHSSSSSYHY 2145

Qy  1721 ELKDGLREQTEKAIL----------DGKVGREEVGVLFQDRNNLRVKS 1758
            ||||||||||||||           |||||||||||||||||||||||
Db  2146 QSEGTERQKGQSGLVWRHGSYGSADYDYGESGFRHSQHGSVSYNSNPVLFKERSDICKAS 2205
```

```
Qy    1759 V----------------SVSQSVSKSEGFN 1772
Db    2206 AFGKDHPRYYATYINKDPGLCGHSSDISKQLGFS 2239


RESULT   5
B82519
hemagglutinin-like secreted protein XF2775 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: B82519
R;anonymous,
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: B82519
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3455 <SIM>
A;Cross-references: GB:AE004082; GB:AE003849; NID:g9108003; PIDN:AAF85560.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.;
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silvei
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF2775


Query Match             2.7%;  Score 255.5;  DB 2;  Length 3455;
Best Local Similarity   18.6%;  Pred. No. 0.00048;
Matches   412;  Conservative 285;  Mismatches 825;  Indels 695;  Gaps 101;


Db    1684 EKYR----DATRTTQQDILNPDAGAPALIQTG--
                :|: |  :||   ||
Qy                                        GAMMIQTDTLRNHYADLLAGGD 1733

Qy     452 DKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSY 511
                 :|| ||  :||
Db                                      KRVLLIDNRALQ-LS----RTDTFHNI 1776

Qy       3 LKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQPST 62
             |||  || :|: :|
Db    1374 IDNLGTGRLYGDHIAL----HAQTLTNRDETSDGHTHAATIAARQRLDIGADTLRNTAN 1428

Qy    1734 QTIVGLPHPTKEKSDDEHKY---KRVLLIDNRALQ-LS----RTDTFHNI 1776

Qy      63 A---ADGISAAHQQKKSFSLRGCL--GTKKF-SRSAPQGQPGTTHSKGATLRDLLARD 114
            | ||::  |  |   |||
Db    1429 AMILSDGDAAI----GATLDNALHATGTATLLDNRSATIDITGTLNITTTTLNNI--RD 1481

Qy     512 SVDQRGQVA-------ILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHG 561
            |:|: :|::   |  |||
Db    1777 STTYRGKVSVEPWSNESRTPTTQIGGRITSGGHQHIAAQTLNNVTDSTHAPEIQHLTYN 1836

Qy     115 DGETQHEAAAPDA--ARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNF 171
             ||   |  :|  |||
Db    1482 N---VHIAHAPDVVTEARMEQPHW--RKNQPN-------GGSGNFRFTSNYDAHDI----- 1525

Qy     562 KSELEAQSVAISHGRLVADSEGKLFSAAIPKQGDGNELRMKKAMPQHALDEHFGHDHQIS 621
            :   :|| ||
Db    1837 PS--TQTLSVVNGVITVTDTSPSLHTVSLADNGPSAGQELTYIPDQSITTP----- 1885

Qy    1482 N---VHIAHAPDVVTEARMEQPHW--RKNQPN-------GGSGNFRFTSNYDAHDI-----A 219

Qy     622 GFFHDDHGQLNALVK--------NN--FRQQHACPLGNDHQFHPGWN 658
            ||      | |||
Db    1886 NAPIRDPAAPPAVIVTPTGPLTLPNNSLFTLHPDAATLITTDPRFTLGRP 1935

Qy     172 GQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSK---------A 219

Qy     659 LTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLT-KGWTGAESDCKQ 716
                   :|   :|| :|||
Db    1936 YTSADTQLHALGDHDT--LHKRLGDGYYEQRLIRE---QLAQLTGRRRLDGYTDDDQQ 1988

Db    1526 ----------YYILNPADIIKDD------PYITPDGQQIHRAIVRLTPQTSAYFYARGLYA 1570

Qy     717 LKKGLDGAAYLLKDGEVK-----RLNINQSTSSI----KHGTENVFSLPHVRN 760
             :||  :|  | |   |||
Db    1989 YRALLDAGVTVAKQHQLRPGIALSADQLAQLTSDIVWLVQQDVQLPDGTTTRALVPRLYL 2048

Qy     220 TTAHADRVEIAQEDDDSEFQQLHQQRLARERENPQQPPKLGVATPISARFQPKLTAVAES 279

Qy     761 KPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREG 820
              :|||:|| | | |||
Db    2049 RPRTGD----LTPDGALLAAASTTINAH-TFNTGTIDARHL---IDINAHTMDQQG 2097

Qy    1571 SQAERRRMDLTARTGDSVVLYYTDR------QDKQPNDHVAAA-----------ATNHS 1613

Qy     821 ---ISGELKDIHVDHKQNLYALTH-EGEVFHQP--REAWQNGAESSWHKL- 865
               |  |:|   |||
Db    2098 GRLTADAIHITTGDFTTLGGQFKARGYLKVHAQGNFLASSTLRDATTQGTRHHSVTELD 2157

Qy     280 VLEGTDTTQSPLKPQSMLKGSGAGVTPLAVTLDKGLQLAPDNPPALNTLLKQTLGKDTQ 339

Qy     866 ---ALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGT- 917
                :||   || :|||
Db    2158 QQAGFTVTGPGAYLGL-STDQAMTHNGVAINNTGTD----GYTSLNA-TGPLHLGTL 2208

Db    1614 AFIGLDTPQQNERFQT------VPITYAPGDRLTYDSNYG------TCTDDCV 1655

Qy     918 SGSQTVFNRLMQGVKGKVIPGSGLTVKLSA---- 947
             ||  :|||
Db    2209 TTHRSDTTQWDPRNSRHSRIDTEYGTSITG-NGDIQLNSGQDINLRAATLHSTQGTITAL 2267

Qy     340 HYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDG 399

Qy     948 QTGGMT---GAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAY 988
            :||:|  |||
Db    2268 ATGNVTITHGDTIQYTSQDSHTKRSGLLNSRTTTHADQQQTQAIGSTLSADKVFVKGNN 2327

Db    1656 RLVTWHDYTD-PDHTLID------MHRG-PNDVR--------DN 1683

Qy     989 ATQHGWQGREGLKPLYEMQGALLKOLDAHNVR---HNAPQPDLQSKLETLDLGEHGAEL 1044
             |: |  :|  |||
Db    2328 ITVTGSHVVSDAGTYMQAEHDLTLQAATHTTQSTYSHHTKQRGL------IRNGGASL 2379

Qy     400 KSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLH--------D 451

Qy    1045 -LNDMKRFRDELEQSATRSVTVLGQHQG-------VLKSNGEINSEFKPSP 1087
             |:    | :|  |||
Db    2380 TLGNQSQRTDSTTATTTTGSLIGATNGNVTLLAGGHYQQIGSDVLSPHGDIDIHAKKVD 2439

Qy    1088 GKALVQSFNVNRSGQDL-----SKSLQQAVHATPPSAESKLQSMLGH 1129
            :     || :|||
Db    2440 --IIQAHHTSQTTQHTATROSGLTVALSTPLIAGAQTAQQMQHAAARSGDPRLQALAGL 2496

Qy    1130 FVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLLLDTVTIGELHELADKAKLVSDHKPD 1189
             :|  |  || :|||
Db    2497 TTALGA------KNTIDAVRQ-DPRALGGLNAS----LTVG 2526

Qy    1190 ADQIKQLRQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVN 1249
             |:  :|  || |||
Db    2527 -----RSTHDSTTTTTSTAAGSNVTAGGNVH---ISATGDGTASTLTIQGSGVRG- 2574

Qy    1250 LTTRTVLESQGSAEFLAKKLKNTLLSLDSGESMSFSRSYG------GGVSTVFVPTLS 1300
             :|   || :|||
Db    2575 -DTMTYLKADGDIALL-AAQNTV---TNQRDNRGRSAGVGVAVNLGSGGTSAGLTAHAS 2628

Qy    1301 KKVPVPVIPGAGITLDRAYNLSFSRTSGG-LNVSFGRD------GGVSGNI 1344
            :||  :|||
Db    2629                                              -D 451
```

Rodrigues, V.; Rosa, A.J.; de Sa, R.G.; Santelli, R.V.; Sa...
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; da Si...
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A...
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2196

Query Match
Best Local Similarity   2.6%; Score 247; DB 2; Length 3442;
                         18.5%; Pred. No. 0.0013;
Matches 412; Conservative 292; Mismatches 815; Indels 708; Gaps 103;

Qy   3 LKSLGTEHKAAVHTAAHNPVGHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQPST 62
         : :||| : : :: ||| : ||| : |: ||
Db 1374 IDNLGTGRLYGDHIAL------HAQTLTNRDETSDGHTHAATIAARQRLDIGADTLRNTAN 1428

Qy   63 A---ADGISAAHQQKKSFSLRGCL--GTKKF-SRSAPQGQPGTTHSKGATLRDLLARD 114
         | ::| | | ||:| |:|
Db 1429 AMILSDGDAAI-----GATLDNALHATGTATLLDNRSATIDITGTLNITTTLNNI--RD 1481

Qy 115 DGETQHEAAAPDA---ARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNF 171
         :|| :| :   ::| :| | : |: :
Db 1482 N---VHIAHAPDVVTEARMEQPHW-RKNQPN------GGSGNFFFTSNYDAHDI-- 1525

Qy 172 GQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSK--------A 219
         :| | :   : | |
Db 1526 ------------YLNPADIIKDD-----PYITPDGQQIHRAIVRLTPQTSAYFYARGGLYA 1570

Qy 220 TTAHADRVEIAQEDDDSEFQOLHQQRLARERENPQQPPKLGVATPISARFQPKLTAVAES 279
              : | : | :: | :
Db 1571 SQAERRRMDLTARTGDSVVLYYTDR------QDKQPNDHVAAA------ATNHS 1613

Qy 280 VLECTDTQSPLKPQSMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQ 339
         |: : | | :   :: :| |
Db 1614 AFIGLDTPQQNERFQT------VPITYAPGDDRLTYDSNYG------TCTDDCV 1655

Qy 340 HYLAHHASSDGSQHLLLDNKGHLEFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSVDG 399
           : : : : || |:| | |::| :: |
Db 1656 RLVTWHDYTD-PDHTLID------MHRG-PNDVR------DN 1683

Qy 400 KSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWOHPAGAARPQGESIRLH------D 451
          || | || ::| : : |:||
Db 1684 EKYR---DATRRTTQQDILNPDAGAPALIQTG-----GAMMIQTDTLRNHYADLLAGGD 1733

Qy 452 DKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSY 511
         : || : | : | :| || :: | :
Db 1734 QTIVGLPHPTKEKSDDEHKY-----------KRVLLIDNRALQ-LS------RTDTFHNI 1776

Qy 512 SVDQRGQVA--------ILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHG 561
         | || |:| | : |: :: | : | | |||
Db 1777 STTYRGKVSEPWSNESRTTPTTQIGGRITSGGHQHIAAQTLNNVTDSTHAPEPIQHLTYN 1836

Qy 562 KSELEAQSVAISHGRLVVVADSEGKLFSAAIPKQGDNELKMKAMPQHALDEHFGHDHQIS 621
         || : |:: | :   ::   :| : | | |:
Db 1837 PS--TQTLSVVNGVITVTDTSPSLHTVSLADNGPSAGQELTYIPDQSITTP----- 1885

Qy 622 GFFHDDHGQLNALVK--------------NN----FRQQHACPLGNDHQFHPGWN 658
                  |: :   : |
Db 1886 ----------NAPIRDPAAPPAVTVTPTGPLTLPNNSLFTLHPDAATLITTDPRFTLGRP 1935

Qy 659 LTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLT--KGWTGAESDCKQ 716
         |: || : | :||| :   : | :||: || : : | :|:
Db 1936 YTSADTQLHALGDHDT--LHKRLGDGYYEQRLIRE----QLAQLTGRRRLDGYTDDDQQ 1988

Qy 717 LKKGLDGAAYLLKDGEVK----------RLNINQSTSSI 760
         |: | || ||||| | |: | | :||
Db 1989 YRALLDAGVTVAKQHLRPGIALSADQLAQLTSDIVWLVQQDVQLPDGTTTRALVPRLYL 2048

Qy 761 KPEPGDALQLNKDDKAQAMAVIGVNKYLALTEKGDIRSFOIKPGTOQLERPAQTLSREG 820
         | || | || ||: | ::: | |: || :| | || | : :| |
Db 2049 RPRTGD---LTPDGALLAAASTTINAH-TFTNTGTIDARHL-----IDINAHTMDQQG 2097

Db 2629 TST------GSGHSTDLTW--SNSHVGGNLLAIDAGGDLLMKGAIGTAKHVIADIAGNL 2680
              ||| | : |: |: | | | | |: :
Qy 1345 MVATGHDVMPYMTGKKTSAGNASDWLSAKHIKISPDLRIGAAVSGTLQGTLQNSLFKLTE 1404

Db 2681 RIESLQDTHQYRSTDRSLGGS--------LTAGAGFSGGSANLNHQTIRSDYASV 2726
              | : :   :   : | | | | : |
Qy 2727 TEQSGLFTCDGGYQLTVGGQTHLIGGAITSNSTAIHNGL-NSLDTGTLLIQNIENHANYT 2785

Qy 1405 DELPGF---------IHGLTH--GTLTP-AELLQKGIEHQMKQGSKLITFSVDTSAN- 1448
         || :: || | | | :| | |:   :| :| :| || ||
Db 1449 ---LDLRAGINLNE----------------DGSKPNGVTARVS------1472

Qy 2786 ATQVNLGGGYSRNGGTVGTDQQGHAATATQVPGTTLPSHNGLSASPPSAMTARDSHSHSTT 2845
           | : | | |: | || | || | | | | | | |
Db 1473 -AGLSASANLAAGSRERSTTSGQFGSTTSASNNRP--------1506

Qy 2846 YSGISQGALTIRDDTAQHALTGHTAAETIATLNRDLLTDTTATSNALTPIFDEQRINAGFD 2905
            |: |: ||| |: | :| | | :: | :| |
Db 1507 --------TFLNGVGAGANL----TAALGVAHSSTHEGKPVGIFPAFTSTNVSAALA 1551

Qy 2906 IVSGLQRETGTFINNRAAEADLKTRQATAADHAAHDPSN----GFNDQQRQTLRDQAIA 2960
         || | : | : | | | : | | |: || |:|||| :
Db 1552 LDNR------TSQSISLELKRAEPVTSNDISELTSTLGKHF------KDSATT 1592

Db 2961 LTNEAHALKDAWGPGGTYRQITTALAAG--ASGNVSAASSDLAKHMIVNYVQQQGATAI 3017
          :: | :| | ||| | : | |:: | : ||: | || :||
Qy 1593 KMLAALKELDDAKPAE-QLHIL----QOHFSA------------KDV 1622

Qy 3018 GHWVATGQLTEGSPLHAALHALLACAGAAASQHCSSGAQGAAASSVLTGLFSDDPREDT 3077
           | || :: ||| ||| || | || | : : || | | |:|
Db 1623 VGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGI-----VELLHK 1677

Db 3078 AQDR--EAKRNLITSIVTGIASIGHT-DPATATHAA----IAAVDNNWLAAKQYVQMLNE 3130
         | | | | |::: | | | | | | ||| | |: | |
Qy 1678 HFDAALPASSAKRLGEMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDG 1737

Db 3131 EFEAA-----TEKEKGRL------EEEKVRAKWRE-IDARQDKLTVDG 3166
          | | || | |:| | ||| | :|
Qy 1738 KV-GREEVGVLFQDRNNLR--VKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIG 1793

Db 3167 LLKGLKESGI-NDINGLEHLFLHPVDTVHELGKI-LTHPTLLLQLGESAVQELLNKVS 3222
           | | | :| ||||: | ||| ||: || | | |||:|||| | ||||:
Qy 1794 TIN--FKYGQDQNTPR------RFTL-EGGIAQANPQVASALTDLKKEGLE 1835

Db 3223 RMSEALLVGGDQHAQQLGEDLGSVIADVGFALAAAGTAKAAEIIGEAGISLSKDVLE 3279
          : | |||| |:| | | |||| |||| | || | || | ||| |||
Qy 3167 ...

RESULT   6
E82589
hemagglutinin-like secreted protein XF2196 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: E82589
R:anonymous
   The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequer
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82589
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3442 <SIM>
A:Cross-references: GB:AE004032; GB:AE003849; NID:g9107324; PIDN:AAF84995.1; GSPDB:GN001
C:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, V.C.A.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

```
Qy    821 ------ISGELKDIHVDHKQNLYALTH-EGEVFHQP--REAWQNGAESSWHKL-  865
                 :|: |  |||||                            ::  ||
Db   2098 GRLTADAIHITTGDFTTLGCQFKARGYLKVHAQGNFLASSTLRDATTQGTRHSVTELD 2157

Qy    866 -ALPQSESKLKSLDMSHEHKPIATFEDGSQHOLKAGGWHAVAAPERGPLAVGT-  917
           ::|    ||  :| :   |  || |||                ||  |||
Db   2158 QQAGFTVTGPGAYLGL-STDQAMTHNGVAINNTGTD------GYTSLNA--TGPLHLGTL 2208

Qy    918 ------SGSQTVFNRLMQGVKGKVIPGSGLTVKLSA-------  947
                    ||:| :    |   |  | ||   |||||
Db   2209 TTHRSDTTQWDPRNSRHSRIDTEYGTSITG-NGDIQLNSGQDINLRAATLHSTQGTITAL 2267

Qy    948 QTGGMT------GAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAY  988
               ::| ||          |   |   : ||     :   |  :
Db   2268 ATGNVTITHGDTIQYTSQDSHTKRSGLLNSRITTTHADQQQTQAIGSTLSADRKVFVKGNN 2327

Db   2328 ITVTGSHVVSDAGTYMQAEHDLTLQAATHTTQSTYSHHTKQRGL-------IRNGGASL 2379

Qy   1045 -LNDMKRFRDELEQSATRSVTVLGQHQG------VLKSNGEINSEFKFSP 1087
            :|:            |    |   :          ||:||          :
Db   2380 TLGNQSQRTDSTTTATTTTGSLLIGATNGNVTLLAGGHYQOIGSDVLSPHGDIDIHAKKVD 2439

Qy   1088 GKALVQSFNVNRSGQDL------SKSLQQAVHATPPSAESKLQSMLGH 1129
             |:   |      ||        |:||| ||                ||
Db   2440 ---IIQAHHTSQTTQHTATRQSGLTVALSTPLIAGAQTAQQMQHAAARSGDPRLQALAGL 2496

Qy   1130 FVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPD 1189
               | |  | :|      ||  ::                  |           ||
Db   2497 TTALGA-----KNTIDAVRQ-DPRALGGLNAS------LTVG 2526

Qy   1190 ADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVN 1249
                 ||  |          |       | |:  |                  :  | ||
Db   2527 --RSTHDSTTTTTTSTAAGSNVTAGGNVH---ISATGDTASTLTIQGSDVRG-- 2574

Qy   1250 LTTRTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYG------GGVSTVFVPTLS 1300
            :|:                         |   ||  ||        ||:| |
Db   2575 -DTMTYLKADGDIALL-AAQNTV----TNQRDNRGRSAGVGVAVNLGSGSTAGLTAHAS 2628

Qy   1301 KKVPVPVIPGAGITLDRAYNLSFSRTSGG-LNVSFGRD------GGVSGNI 1344
                     ||   ||           ||          |       ||:|
Db   2629 TST------GSGHSTDLIW--SNSHVGGNLLAIDAGGDLLMKGAIGTAKHVIADIAGNL 2680

Qy   1345 MVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTE 1404
                 |                  |                  |  ||: :|  :  ||
Db   2681 RIESLQDTHQYRSTDRSLGGS------LTAGAGFSGSGANLNHQTIRSDYASV 2726

Qy   1405 DELPGF------IHGLTH--GTLTP-AELLQKGIEHQMKQGSKLTFSVDTSAN- 1448
            |   |         ||    |||  |    || |         |||||
Db   2727 TEQSGLFTGDGGYGLTVGGQTHLIGGAITSNSTAIHNGL-NSLDFGTLILQNIENHANYT 2785

Qy   1449 --LDLRAGINLNE------DGSKPNGVTARVS------ 1472
              :|  |           |||||| ||||||
Db   2786 ATQVNLGGGYSRNGGTVGTDQQGHAATATQVPGTTLPSHNGLSASPPSAMTARDSSHSTT 2845

Qy   1473 -AGLSASANLAAGSRERSTTSGQFGSTTSASNNRP------ 1506
             | |          ||||  ||                 |
Db   2846 YSGISQGALTIRDDTAQHALTGHTAAETIATLNRDLLTDFTATSNALTPIFDEQRINAGFD 2905

Qy   1507 ------TFLNGVGAGANL----TAALGVAH------SSTHEG 1532
                   ||||||||||      ||||                ||
Db   2906 IVSGLQRETGTFINNRAAEADLKTRQATAADHAAHDPSNGFNDQQRHALREQAIALTNEA 2965

Qy   1533 KPV--GIFPAFTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTLGKHF--- 1586
            |||   ||                     ||||||||||||
Db   2966 HAIKDAWGPGGTYRQITTALAAG--ASGNVS------AASSNVSAASSDLAKHMIVNY 3015

Qy   1587 --KDSATTKMLAALKELDDAKPAE-QLHIL------QQHFSA- 1619
              :|      ||                         |  :|
Db   3016 VQQQGATAIGHWVATGQLTEGSPLHAALHALLACAGAAASQHCSSCAQGAAAASSVLTGL 3075

Qy   1620 ------KDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGI- 1671
                   |  |                    |                |||||||:|
```

```
Db   3076 FSDPRPEDTTQDR--EAKRNLITSIVTGIASTGNT-DAATATHAA----IAAVDNNWLAA 3128
            |  |            |||||:||||                      |||||||||||
Qy   1672 ---VELLHKHFDAALPASSAKRIGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLRE 1728
               ::|:||||                              ::|  ::  |  :||  :
Db   3129 KQYVQMLNEEFEAA--TEKEKGRL                                     3164

Qy   1729 QTEKAILDGKV-GREEVGVLFQDRNNLR--VKSVSVSQSVSKSEGFNTPALLLGTSNSA 1784
            : ||   |    | :  ||                          |||||||||
Db   3165 RQDKLTVDGLLKGLKLKESGI--NDINGLEHLFLHPVDTVHELGKI--LTHPTLLLQLGESA 3220

Qy   1785 AMSMERNIGTIN-FKYGQDQNTPRRFTLEGG--IAQANPQVASALT------D 1828
            |   :|         :| |||                :|          |
Db   3221 VQELLNKVSRMSEALLVGGDQHA-QQFGEDLGSVIADVGVALAAAGTFKAAEILGEAGIN 3279

Qy   1829 LKKEGLE 1835
            |||: |:
Db   3280 LSKDVLE 3286
```

RESULT 7

H81193

hemagglutinin/hemolysin-related protein NMB0493 [imported] - Neisseria meningitidis (

C;Species: Neisseria meningitidis

C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Dec-2000

C;Accession: H81193

R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000

A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.

A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A;Reference number: A81000; MUID:20175755

A;Accession: H81193

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-2703 <TET>

A;Cross-references: GB:AE002405; GB:AE002098; NID:g7225708; PIDN:AAF40927.1; PID:g72

A;Experimental source: serogroup B, strain MMD58

C;Genetics:

A;Gene: NMB0493

Query Match   2.6%; Score 245; DB 2; Length 2703;
Best Local Similarity 19.0%; Pred. No. 0.0011;
Matches 414; Conservative 282; Mismatches 766; Indels 712; Gaps 105;

```
Qy    10 HKAAVHTAAHNPVGHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQPSTAAD---G  66
          | |||:||||||||||:                 ||                     |
Db   573 HDLAVNTQTAKNSGH-LLTQTGKIDNRELHNAGEIAA----NNLTLIHSGRLSNDKKGN 626

Qy    67 ISAAHQQKKSFSLR------GCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGE 117
          |  ||                                            ||  ||||||
Db   627 IRAAHLQLDTAGLHNAGNILADSGVTVTTKNNLRNTGKVSVARLNTEGQTLDNTRGRIEAE 686

Qy   118 TQHEAAAPDAARLTRSGG------VKRRNMDDMAGRPM------VKGG-----SGE 156
           :       |   |  ||       |:|                  ||||      ||
Db   687 TVN----IQSQLTNQSGHITATEQLTINSRNVDNQNGKLLSANQAOLAVSDGLYNQHGE 742

Qy   157 DKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGST 216
              |  |  ||                         ||               |  ||
Db   743 IATNRQLSIHDKN----QNTLALNNADGTIOSAGN------VSLQAKSLA 782

Qy   217 SKATTAHADRVEIAQEDDDSEFQQLHQQRLARERENPPQPPKLGVATPISARFQPKLTAV 276
                 |||    |                                           ||
Db   783 NNGTLTTAGNKLDIALTDD------FVVERD------LTAGKQLNLS- 816

Qy   277 AESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGK 336
           :  |||  :          ||              |||:  |         ||:||:||
Db   817 IKGRLKNTHTLQ------AGHT--LKLNAGNI---DN----QVTGK 847

Qy   337 DTQHYLAHHASSDGSQHLLLDKGHLLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVS 396
            | |||
```

```
Db  848 -----IIGGEQTDITSEQHVDNRG----------------LINS-------DGLTHIGAGQTL  882
                                                              :: |  ||:||||  |:
Qy  397 VDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGE--SIRLHDDKI  454
         |  :  | ::  | : ||  | | : :  :||| |     ::| |  || | ||
Db  883 TNTGTGKI-YGNHIALDAQILLNREETTEGSTKAG-----AIAARKRLDIGAKEIHNQEG  936
Qy  455 HILHPE-LGVWQSADKDTHSQ-----------LSRQADGKLYALKDNRTLQNLSDNKS  500
         ||  | |  :  :||  |||          ||::| |  | : :|  |:|:::::
Db  937 ALLSSEGIFAVGNRLDEQHHAAGMADTFVNGSAGLEVQGD----ALMSVRNMQNINNHFK  992
Qy  501 SE----KLVDKIKSYSV-----DQRGQVALLTDTPGRHKMSIMPSLDASPESHISLS-L  549
         |:    ||: : :| |      ||| |  | || | : ||  | | ::| ||:| | |
Db  993 TETYLAKAEKQVRDYTVLGQNTYYQAGKDGLFDNSQGQ-KDQTTATFHLKNGSRIEANQW  1051
Qy  550 HFADAH-----QGLLH-----------GKSELEAQSVAISHGRLVVAD-------  581
         | ||:      ||||             |||| ||| :| | | | |:
Db 1052 HVRDYHIETYKERIIENRPAHITVGDLTASGQNWLNKDSRIVVGGRIITDDLNQKEITN  1111
Qy  582 --SEGKLFSAAIPKQGD-----------GNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQ  630
           ::| ||::| |:|:            | : | :| ||| || :| | || |::| | |
Db 1112 QSTTCKGRTDAVGTQWDSVTKKGWTYSGRKRQRRTERNHTP---YHDTQL--FTHDFDTP  1165
Qy  631 LNALVKN----NFRQQHAC-----PLGNDHQFHPG--WNLTDALVIDNQLGLHHTN  675
         |:|  :     | |:| |      ||| | || |   | || | | |  | | ||
Db 1166 VSVIQQNAASPSFQPAASAIKLIDGVSTAAVNGQRIHTGNVVSLNNATVTLPNSSLYTTH  1225
Qy  676 PEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKR  735
         || ||| |:|||| ||:|||:|| ||  | | | ||: || : |:|||| ||| |:| |
Db 1226 PD----NKGW-----LVETDPQFADY--RRWLGSDYMLQOLQLDTNHLHKRLGDGYYEQ  1273
Qy  736 LNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDKAQAMAVIGVNKYLALTEKG  795
          : :::| | ||| ||  : |          :| || ||  | ||: |:|:   : ||
Db 1274 KLVN------EQIHQLTGYRR-----LDGYRSDEE--------QFKALMDNG  1306
Qy  796 --DIRSFQIKPG----TQQ-------LERPAQTLSREGISGELKDIHVDHKQNLYAL  839
           ||:| | ||    |||       | | | || ||:|  | |:|:||||:|: ||
Db 1307 LTAAKTFGLTPGIALSAEQVARLTSDIVWMENQTVTLS---DGSTQTVLV--PKVYAL  1359
Qy  840 THEGEV------FHQPREAWQNGAESSS---WHKLALPQSES-------  872
         || :       | ||  |  || :|     | ||| |:||
Db 1360 ARKGDLNTSGGLISAEQVLLKLQNGNLTNSGTIAGRQAVLIQARNINSNGNIQADQIGLK  1419
Qy  873 KLKSLDMSHEHKPIATFEDGSQHQ----------LKAGGWHAYAAPER-------G  911
         |:| ||  :||  : | ||| |            ||| | : | ||          |
Db 1420 AEKSINI------DGGQVQAGRLLTAQAQNINLNCTTQTSGNERNGNTAIDRMAG  1468
Qy  912 PLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGMTGAEGRKVSSKFSE--RI  968
          | ||| |:||  |||:|||:|||:||||||  |:  || | || |:||||||   |
Db 1469 INVVG-SHTEQVDNRTSDGI----------LSLHASNDINLNAATVSNQVKDGTTQI  1514
Qy  969 RAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDL  1028
         :|       | | | |:::        :|  || | |:| :|:|| || | || :|
Db 1515 TA-GNNLNLGTIRTEHREAYGT-------LDDENHRHVRQSTEV  1550
Qy 1029 QSKLETLDLGEHGAELL--NDMKRFRDELEQ-------SATRSVT------VL  1066
         ||:| || | |  : |   |    | |:||        |  |:|      ||
Db 1551 GSSIRT----QNGALLRAGNDLKIRQGELEAEEGKTVLAAGRDVTISEGRQITELDTSVS  1606
Qy 1067 GQHQGVLKSNGEINSEFKPS----------PGKALV--QSFNVN-------RSG  1101
         |  :| |:|:  :: || |           || |:  |||:|         | |
Db 1607 GKSKGILSST-KTHDRYRFSHDEAVGSNIGGGKMIVAAGQDINVRGSNLISDKGIVLKAG  1665
Qy 1102 QDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHOKGEIPLGRQR--DPNDKTALT  1159
          | | ||:::| || | ||| | | || | || | | |||| ||: |    |: || |
Db 1666 HDIDISTAHNRYTGNEYHESKKSGVMG----TGGLGFT-------IGNRKTTDDTDRTNIV  1715
Qy 1160 KSRLIL-----DTVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVK  1214
          |: : |    || | || || | ||: | |: | || :| |  :  :| :|:|::|
Db 1716 HTGSIIGSLNGDTVTVAGNRYRQTGSTVSSPEGRNTVTAKSIDVEF---ANNRYATDYAH  1772
Qy 1215 HYTDMGFT--HNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNT  1271
         | || | |  || ||:| ||||  | ||| :|| |::|| ||| ||| |:|| |||:|
Db 1773 TQEQKGLTVALNVPVVQAAQNFIQAAQNVGKSKNKRVNAMAAANAAWQ-SYQATQOMQQF  1831
```

```
Qy 1272 LLSLDSGESMSFSRSYGGGVSTVFVPTLSK-------------KVPVPVIPGAGITLDRAYNLSF  1323
         || |  |||| | | | |||  |  |  | |             ||| || | | ||| : || |
Db 1832 APSSSAGGQNNNQSPSISVSITYGEQKSRNEQKRHYTEAAASQIIGKGQT-----------TL  1884
Qy 1324 SRTSGG-----LNVSFGRDGGVSGNIMVATGHDVM--PYMTGKKTSAGNASDW--LSAKHKI  1376
         | |:|      ||:  |:| |:|||: ||| | |   | |||| | | | ||  ||:|: |
Db 1885 AATGSGEOSNINITGSDVIGHAGTALIADNHIRLQSAKQDGSEQSKNKSSGWNAGVAVKI  1944
Qy 1377 SPDLRIGAAVSGTL-QGTLQNSLKFKLTEDELPGFIHGLTH------------GTLT  1420
         |  :||| |: |:: ||:: |::|   ||:   |     | |            |||
Db 1945 GNGIRFGITAGGNIGKGKEQGG-------------STTHRHTHVGSTTGKTTIRSGGDTTLK  1993
Qy 1421 PAELLQKGIE----------------HQMKQ---------------GSKLT-  1440
         | :| :|||                 ||:||               |||:
Db 1994 GVQLIGKGIQADTRNLHIESVQDTETYQSKQQNGNVQVTVGYGFSASGSYRQSKVKADHA  2053
Qy 1441 --------------FSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSR  1486
                       |||: | | | | ||:| |:| | |||| ||||:|| |: ||: |
Db 2054 SVTGQSGIYAGEDGYQIKVRDNTDLKGGIITSSOSAEDKGKNLFQTATLTAS-DIQNHSR  2112
Qy 1487 ERSTTSGQFGS-------------TTSASNNRPTFLNGVGAG-------ANLTAALGV-----AH  1526
         | | | | ||               |||| |||| |||| |        |:| | ||     ||
Db 2113 YEGRSFGIGGSFDLNGGWDGTVTDKQGRPTDRISPAAGYGSDGDSKNSTTRSGVNTHNIH  2172
Qy 1527 SSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSISLELKRA-EPVTSNDIS---ELTST  1581
         ||: :|| | || |  |||| | | ||||:||::| | |:  || | ||    |||:
Db 2173 ITDEAGQLARTGRTAKETEARIYTGIDTETADQHSGHLKNSFDKDAVAKEINLQREVTKE  2232
Qy 1582 LGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQ  1641
         | | | || | |  |||: |: ||| ||| :| :|| ||: |||||| ::||  |:| |
Db 2233 FGRN-----AAQAVAAV------------ADKLGNTQSY----------ERYQEARTL-LEAEL  2268
Qy 1642 QAADSHSME--------------------LGSASHSTTYNNLSRINN  1668
         | | |::|                     || | | |:: : |:||
Db 2269 QNTDSEAEKAAFRASLGQVNAYLAENQSRYDTWKEGGIGRSILHGAAGGLTTGSLGGILA  2328
Qy 1669 DGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLRE  1728
         :| :  | ||||| |||:||:||||:| || |||: || |: | || | | |||||| |
Db 2329 GGGTSLAAPYLDKA-----AENLG-----PAGKAAVNALGGAAIGYAT------------  2366
Qy 1729 QTERAILDGKVGREEVGVLFQDRNN---LRVKSVSVS----QSVSKSEGFNTPALLL  1778
         |:| | |:| | ||:|||:|:|||    ||:||||:    ||:||| | ||| |||
Db 2367 -------------GGSGGAVVGANV-DWNNRQLHPKEMALADKYAEALKREVEKREGRKI------------  2412
Qy 1779 GTSNSAAMSMERNI  1792
         |||||| | :| :
Db 2413 -SSQEAAMRIRRQI  2425
```

RESULT 8

A56539

giantin - human

N;Alternate names: macrogolgin
C;Species: Homo sapiens (man)
C;Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999
C;Accession: A56539; S37536
R;Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
A;Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane prote
Mol. Cell. Biol. 14, 2564-2576, 1994
A;Reference number: A56539; MUID:9187728
A;Accession: A56539
A;Molecule type: mRNA
A;Residues: 1-3259 <SEE>
A;Cross-references: EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g405715
C;Genetics:
A;Gene: GDB:GOLGB1; GCP; GCP371
A;Cross-references: GDB:454958
A;Map position: 3q13.31-3q13.31
C;Superfamily: giantin
C;Keywords: coiled coil; Golgi apparatus; transmembrane
F;3238-3254/Domain: transmembrane #status predicted <TMN>

```
Query Match      2.6%;     Score 244.5;  DB 1;  Length 3259;
Best Local Similarity  18.6%;  Pred. No. 0.0016;
Matches  389;  Conservative 311;  Mismatches 816;  Indels 577;  Gaps 92;
```

Left column:

```
Qy    23 GHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQ--PSTAADGISAAHQQKKSFSLR   80
Db   603 GEGIA--------PIKMVFLEDTGQDFPLMPNEESSLPAVEKEQASTEHQSRTSEEI-   652

Qy    81 GCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQHEAAAPDAAR---------  129
Db   653 --------SLNDAGVELKS--TKQGD-KSLSAVPDIGQCHQDELERLKS          691

Qy   130 LTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNN------FGQMR        175
Db   692 QILELELNFHKAQEIYEKNLDEKA-----KEISNLNQLIEEFKKNADNNSSAFTALSEER 746

Db   747 DQLLSQVKELSMVTELRAQVKQLEMNLAEAERQRRLDYES---QTAHDNLLTEQIHSLSI 803

Qy   176 QTMLSKM----------------AHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSK  218

Qy   219 ATTAHADRVEIAQED-DDSEFQQLHQQRLARERENPPQPPKLGV-----ATPISARFQP  271
Db   804 EAKSKDVKIEVLQNELDDVQLQFSEQSTLIRSLQSQLQNKESEVLEGAERVRHISSKVEE 863

Qy   272 QSMLKGSGAGVTPLAVTLDKGKLQLAPDN-                              322
Db   864 LSQALSQKELEITKMDQLLLEKKRDVETLQQTIEEKDQQVTEISFSMTEKMVQLNEEKFS 923

Qy   323 PPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSV         374
Db   924 LGVEIKTLKEQLNLLSRAEEEAKKEQVEEDNEVSSGLKQN------YDEMSPAGQISK   974

Qy   375 LHNSHPGEI------GKGKLAQAGTGSVSVVDGKSGKIS-----LGSGTQSHNKTMLSQP-- 421
Db   975 EELQHEFDLLKKENEQRKRKLQAALINRKELLQRVSRLEEELANLKDESKKEIPLSETER 1034

Qy   422 GEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADG  481
Db  1035 GEVEEDKENKEYSEKCVTSKCQEIEIYL--KQTISEKEVEL-QHIRKDLEEKLAAEEQF  1090

Qy   482 KLYALKDNRTLQN------LSDNKSSEKLVDKIKSYSVD--QRGQVAILTDTPGR      528
Db  1091 QALVKQMNQTLQDKTNQIDLLQAEISENQA---IIQKLITSNTDASDGDSVALVKET-  1144

Qy   529 HKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFS  588
Db  1145 --VVISPPCTGSSE----------HWKPELEEKILALEKEK--EQLQKKLQE         1182

Qy   589 A-------AIPKQGD-----GNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN 637
Db  1183 ALTSRKAILKAQEKERHLREELKQQKDDYNRLQEQFDEQSKENENIGDLRQLQIQVRE  1242

Qy   638 NF-----RQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHLGSL   690

Qy   691 ALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAY--LLKDGEVKR---LNINQSTSS  744
Db  1282 --SNLCPDWPSHSEDASALQGGTSVAQIKAQLKEIEAEKVELELKVSSTTSE        1331

Qy   745 IKHGTENVFSLPHVRNKPEPGDALQGLNKDKAQAMAVIGVNKYLALTEKGDIRSFQIKP  804
Db  1332 LTKKSEEVFQLQEQINK-----QGL-----EIES--LKT                    1358

Qy   805 GTQQLERPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFHQPRE---AWQNGAESSS 861
Db  1359 VSHEAEVHAESLQQKLESSQLQIAGLEHLRELQPKLDELQKLISKKEEDVSYLSGQLSEK 1418

Qy   862 WHKLA-----LPQSESKLKSLDMSHEHKPIATFEDGSQHLKAGGWHAYAAPERGPLAVG  916
Db  1419 EAALTKIQTEIIEQEDLIKALHTQLEMQ--AKEHDERIKQLQVELCEMKQKPEE----IG 1472
```

Right column:

```
Qy   917 TSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERI---RAYAF  973
Db  1473 EE---SRAKQQIQRKL-                         QAALISRKFALKENKSLQEELSLARGTIE 1514

Qy   974 NPTMSTPRPIKNAAYATQHGWQGRE------GLKPLYEMQGALIKQLDAHNVRHNAPQPD 1027
Db  1515 RLTKSL------ADVESQVSAQNKEKDTVLGRLALLQEERDKLITEMD-    RSLLENQS 1564

Qy  1028 LQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSP 1087
Db  1565 LSSSCESLKLALEG--LTEDKEKLVKEIESLKSSKIAESTEWQ----EKHKELQKEY---- 1615

Qy  1088 GKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSH----QKG  1142
Db  1616 -EILLQSY---ENVSNEAERIQHVEAVRQER-QELYGKLRSTEANKKETEKQLQEA    1667

Qy  1143 EIPLGRQRDPNDKTALTKSRLILDTV---TIGELHELADKAK----LVSDHKPDADQ   1192
Db  1668 EQEMEMKEKMRKFAKSKQQKILELEEENDRLRAEVHPAGDTAKECMETLLSSNASMKEE 1727

Qy  1193 IKQLRQQFDTLREK------RYESNPVKHYTDMGFTHNKALEANYDAVKAFINAF     1241
Db  1728 LERVKMEYETLSKKFQSLMSEKDSLSEEVQDLKHQIE--DNVSKQANLEAT         1776

Qy  1242 KKEHHGVNLTTRTVLESQGSAELAKKLKNTLLSLDS---GESMSFSRSYGGGVSTVF-  1295
Db  1777 EKHDNQTNVTEEGTQSIPGETE----EQDSLSMSTRPTCSESVPSAKSANPAVSKDFSS 1831

Qy  1296 VPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGG----LNVSFGRDGGV          1340
Db  1832 HDEINNYLQQIDQLKERI-----AGLEEEKQKNKEFSQTLENEKNTLLSQISTKDGEL  1884

Qy  1341 SGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQG        1392
Db  1885 KMLQEVTKMNLLNQQIQEELSRVTKLKETAEEKD------DLEE               1924

Qy  1393 TLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVD--     1444
Db  1925 RLMNQLA------ELNGSI-GNYCQDVTDAQIKNELLESEMKNLKKCVSELEEEKQQLVK 1977

Qy  1445 -TSANLDLR------AGINLNEDGSKPNGV-                 TARVSAGLS    1476
Db  1978 EKTKVESEIRKEYLERKIQGAQ-KEPGNKSHAKELQELLKEKQQEVKLQLQKDCIRYQEKIS 2036

Qy  1477 ASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVG 1536
Db  2037 A-------LERTVKALEFVQTESQKDLEITKEN-          LAQAVEHRKKAQAELAS 2079

Qy  1537 IFPAFTSTNVSAALAL-DN----RTSQSISLELKRAEPVTSNDIS-ELTSTLGKHF    1586
Db  2080 FKVLLDDTQSEAARVLADNLKLKKELQSNKESVKSQMKQD----EDLERRLEQAEEKHL  2135

Qy  1587 KDSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADS 1646
Db  2136 KEKKN------MQEKLDALRREKVHL----EETIGEIQVTLNKKDKEVQQLQENLDS   2182

Qy  1647 HSMELG--SASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMNNDPALKDI  1704
Db  2183 TVTQLAAFTKSMSSLQDDRDRV-IDEAKKWERKFSDAIQSKEEEIRLKE--DNCSVLKDQ 2239

Qy  1705 IKQLQSTPFSSASVSM-ELKDGL-REQTEKAILDGKVGRE------             1742
Db  2240 LRQM--------SIHMEELKINISRLEHDKQIWESKAQTEVQLQQKVCDTLQGENKELLS 2291

Qy  1743 EVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSM            1788
Db  2292 QLEETRHLYHSSQNELAKLESELKSLKDQLTDLSNSLEKCKEQK----GNLEGIIRQQ  2345

Qy  1789 ERNIGTINFKYGQ--DQNTPREFT-LEGGIAQANPQVASALTDLKKEGLEM         1836
Db  2346 EADIQNSKFSYEQLETDLQASRELTSRLHEEINMKEQKIISLLSG--KEEAIQV      2397
```

```
RESULT   9
I52300
giantin - human
N;Alternate names: gcp372
C;Species: Homo sapiens (man)
C;Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C;Accession: I52300
R;Sohda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A;Title: Molecular cloning and sequence analysis of a human 372-kDA protein localized in
A;Reference number: I52300; MUID:95100974
A;Accession: I52300
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3225 <RES>
A;Cross-references: GB:D25542; NID:g662389; PIDN:BAA05025.1; PID:g808869
C;Superfamily: giantin

Query Match         2.6%;   Score 243.5;   DB 2;   Length 3225;
Best Local Similarity  18.6%;   Pred. No. 0.0017;
Matches  389;   Conservative  311;   Mismatches  816;   Indels  577;   Gaps  92;

Qy      23  GHGVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQ--PSTAADGISAAHQQKKSFSLR   80
                ::|:|                 |:|
Db     569  GEGIA------PIKMKVFLEDTGQDFPLMPNEESSLPAVEKEQASTEHQSRTSEEI-    618

Qy      81  GCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQHEAAAPDAAR---------    129
                |||   :|:|  |  |  ||| :|||
Db     619  ------SLNDAGVELKS--TKQGD-KSLSAVPDIGQCHQDELERLKS            657

Qy     130  ---------LTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNN-----FGQMR  175
                            ::|  |   :||||    |:      :||
Db     658  QILELELNFHKAQEIYEKNLDEKA------KEISNLNQLIEEFKKNADNNSAFTALSEER  712

Qy     176  QTMLSKM----------------AHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSK   218
                |:|                    || :| ||| | |  ||  |  |:||  ||
Db     713  DQLLSQVKELSMVTELRAQVKQLEMNLAEAERQRRLDYES---QTAHDNLLTEQIHSLSI  769

Qy     219  ATTAHADRVEIAQED-DDSEFQQLHQQRLARERENPPQPPKLGV-----ATPISARFQP   271
                    |:|  | :  ||:| |    |||  |::|   ||   |      || |||| |
Db     770  EAKSKDVKIEVLQNELDDVQLQFSEQSTLIRSLQSQLQNKESEVLEGAERVRHISSKVEE  829

Qy     272  KLTAVAESVLEGTDTTQSPLKP-------QSMLKGSGAGVTPLAVTLDKGKLQLAPDN--  322
                |||   :| :|| |          |:||| |  | |||  |||  |||:|:||||
Db     830  LSQALSQKELEITKMDQLLLEKKRDVETLQQTIEEKDQQVTEISFSMTEKMVQLNEEKFS  889

Qy     323  ------------PPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSV  374
                            |   | :||  |:| :|      ||  :|||:| |||  ||
Db     890  LGVEIKTLKEQLNLLSRAEEAAKEQVEEDNEVSSGLKQN------YDEMSPAGQISK      940

Qy     375  LHNSHPGEI-------KGKLAQGTGSVSVDGKSGKIS--LGSGTQSHNKTMLSQP--    421
                |:|             | | ||:|  |   | |:||    | ||:    :| |::
Db     941  EELQHEFDLLKKENEQRRKLQAALINRKELLQRVSRLEEELANLKDESKKEIPLSETER  1000

Qy     422  GEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADG   481
                |||       |:   |:|     | |   | :|| ::|| |
Db    1001  GEVEEDKENKEYSEKCVTSKCQEIEIYL--KQTISEKEVEL-QHIRKDLEEKLAAEEQT  1056

Qy     482  KLYALKDNRTLQN-----------LSDNKSSEKLVDKIKSYSVD--QRGQVAILTDTPGR   528
                |||       |              :|:|::    | |||  |   |||  ::|  ||
Db    1057  QALVKQMNQTLQDKTNQILDLLQAEISENQA---IIQKLITSNTDASDGDSVALVKET---  1110

Qy     529  HKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFS   588
                |   :|   |   ||  :| ::  | |  |:|:||  |  | |   ||    :| |:
Db    1111  --VVISPPCTGSSE-------------HWKPELEEKILALEKEK--EQLQKKLQE     1148

Qy     589  A-----AIPKQGD----GNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKN   637
                |       |  |     || |  | |  | | ||  | |  ||  | | || |:|
Db    1149  ALTSRKAILKKAQEKERHLREELKQQKDDYNRLQEQFDEQSKENENIGDLRQLQIQVRE  1208

Qy     638  NF------RQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMGHLGSL   690
                :         |:| ||  | :|   |||
Db     638  ?
```

```
Db    1209  SIDGKLPSTDQQESC----------SSTPG--LEEPLFKATE--QHHTQP-------VLE----  1247
               ::|:|                          |     :|
Qy     691  ALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAY--LLKDGEVKR------LNINQSTSS  744
                           |: |                     ::||       ::|| :
Db    1248  ------SNLCPDWPSHSEDASALQGGTSVAQIKAQLKEIEAEKVELELKVSSTTSE  1297

Qy     745  IKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKP  804
                |       :|                        |
Db    1298  LTKKSEEVFQLQEQINK--------QGL-----------------EIES--LKT  1324

Qy     805  GTQQLERPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFHQPRE---AWQNGAESSS  861
                |:|::|    :  |||  | :||     |  |  | :||  ::|
Db    1325  VSHEAEVHAESLQQKLESSQLQIAGLEHLRELQPKLDELQKLISKKEEDVSYLSQQLSEK  1384

Qy     862  WHKLA----------LPQSESKLKSLDMSHEHKPIATFEDGSQHOLKAGGWHAYAAPERGPLAVG  916
                :|              |:| :| |   ::|    | | | |    | |    | :||
Db    1385  EAALTKIQTEIIEQEDLJKALHTQLEMQ--AKEHDERIKQLQVELCEMKQKPEE------IG  1438

Qy     917  TSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERI---RAYAF  973
                       | |   |:||| ||     | :::|        |  :||| | |
Db    1439  EE--------SRAKQQIQRKL--------------QAALISRKEALKENKSLQEELSLARGTIE  1480

Qy     974  NPTMSTPRPIKNAAYATQHGWQGRE----GLKPLYEMQGALIKQLDAHVRHNAPQPD  1027
                  |:      |  |   |  |              :|  |: :| :  :|   | |
Db    1481  RLTKSL-----ADVESQVSAQNKEKDTVLGRLALLQEERDKLITEMD--RSLLENQS  1530

Qy    1028  LQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSP  1087
                |:||            |     | :|      :|| || : |
Db    1531  LSSSCESLKLALEG--LTEDKEKLVKEIESLKSKIAESTEWQ--EKHKELQKEY----  1581

Qy    1088  GKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSH----QKG  1142
                :||           |   ||          : |       ||  :|
Db    1582  -EILLQSY-----ENVSNEAERIQHVVEAVRQEK-QELYGKLRSTEANKKETEKQLQEA  1633

Qy    1143  EIPLGRQRDPNDKTALTKSRLILDTV------TIGELHELADKAK-----LVSDHKPDADQ  1192
                :|            |  ||  |  |         | :|| |        |: :| |  |
Db    1634  EQEMEMKEKMRKFAKSKQQKILELEEENDRLRAEVHPAGDTAKECMETLLSSNASMKEE  1693

Qy    1193  IKQLRQQFDTLREK--------------RYESNPVKHYTDMGFTHNKALEANYDAVKAFINAF  1241
                | || |   |:|                  | |      |    |:|  |   |   | |
Db    1694  LERVKMEYETLSKKFQSLMSEKDSLSEEVQDLKHQIE--GNVSKQANLEAT------  1742

Qy    1242  KKEHHGVNLTTRTVLESQSGSAELAKKLKNTLLSLDS---GESMSFSRSYGGGVSTVF--  1295
                ||          |        | |||        | |          | | |
Db    1743  EKHDNQTNVTEEGTQSIPGETE-------EQDSLSMSTRPTCSESVPSAKSANPAVSKDFSS  1797

Qy    1296  -------------VPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGG----LNVSFGRDGGV  1340
                             | |   | |  |     :| |   ||  |
Db    1798  HDEINNYLQQIDQLKERI--------AGLEEEKQKNKEFSQTLENEKNTLLSQISTKDGEL  1850

Qy    1341  --------SGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQG  1392
                          || :|  |  | | |         :||      |  |  | |  : |
Db    1851  KMLQEEVTKMNLLNQQIQEELSRVTKLKETAEEEKD----------DLEE  1890

Qy    1393  TLQNSLKFLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKQGSKLTFSVD------  1444
                ||      ||  |     |     |   | | |:       | |     |
Db    1891  RLMNQLA----ELNGSI-GNYCQDVTDAQIKNELLESEMKNLKKCVSELEEEKQQLVK  1943

Qy    1445  --TSANLDLR--------AGINLNEDGSKPNGV-------------TARVSAGLS  1476
                    :|  |          |:|:   |                  |||  | |:
Db    1944  EKTKVESEIRKEYLEKIQGAQ-KEPGNKSHAKELQELLKEKQQEVKQLQKDCIRYQEKIS  2002

Qy    1477  ASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGKPVG  1536
                |  :|          |:                        |          || |
Db    2003  A-----LERTVKALEFVQTESOKDLEITKEN---------LAQAVEHRKKAQAELAS  2045

Qy    1537  IFPAFTSTNVSAALAL-DN---------RTSQSISLELKRAEPVTSNDIS-ELTSTLGKHF  1586
                | |     | | |    |            | :| | | |   |   || | |
Db    2046  FKVLLDDTQSEAARVLADNLKLKLKELQSNKESVKSQMKQD---EDLERRLEQAEEKHL  2101

Qy    1587  KDSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKFLVIRQQAADS  1646
                                              :| |     |      :|   |
Db    2102  KEKKN---------MQEKLDALRREKVHL----------EETIGEIQVTLNKKDKEVQQLENLDS  2148
```

```
Qy   1647  HSMELG--SASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDI  1704
                :   |:  :: |:   :| :             |  |::   |||
Db   2149  TVTQLAAFTKSMSSLQDDRDRV-IDEAKWERKFSDAIQSKEEEIRLKE--DNCSVLKDQ  2205

Qy   1705  IKQLQSTPFSSASVSM-ELKDGL-REQTEKAILDGKVGRE------                1742
                                                                |
Db   2206  LRQM-------SIHMEELKINISRLEHDKQIWESKAQTEVQLQQKVCDTLQGENKELLS  2257

Qy   1743  ----------EVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLGTSNSAAMSM      1788
                     |:    :  :|: :: |:::
Db   2258  QLEETRHLYHSSQNELAKLESELKSLKDQLTDLSNSLECKEQK------GNLEGIIRQQ  2311

Qy   1789  ERNIGTINFKYGQ---DQNTPRRFT-LEGGIAQANPQVASALTDLKEGLEM          1836
            :|:||:        ||| ::     |||||||:|: : :::  |:   :| :::
Db   2312  EADIQNSKFSYEQLETDLQASRELTSRLHEEINMKEQKIISLLSG-KEEAIQV        2363
```

RESULT   10
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N;Alternate names: hypothetical protein EG:49E4.1
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13564
R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17689
A;Accession: T13564
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5327 <SPA>
A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0025392
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A;Note: EG:49E4.1
C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match   2.5%;   Score 240.5; DB 2; Length 5327;
Best Local Similarity 18.4%; Pred. No. 0.0056;
Matches 370; Conservative 271; Mismatches 803; Indels 567; Gaps 81;

```
Qy     18  AHNPVGHGVALQQGSSSSSPQNAAAASLAAEGKNRGKMPRIHQPSTAADGI------SAAH    71
            : : :::: :    :     : ::          :  :::  :
Db   3517  ASRPTSVAESVKDEAEKSKEESRRESVAEKSSLASK--EASRPASVAESVKDEAEKSKEE  3574

Qy     72  QQKKSFSLRGCLGTKKFSRSAPQGQ--PGTTHSKGATLRDLLARDDGETQHEAAAPDAA    128
             :  :::  :          :  :     : ::                :
Db   3575  SRRESVAEKSPLASKEASRPASVAESVKDEAEKSKEVSRRESVAEKSPLPSKEASRPTSV  3634

Qy    129  ----------RLTRSGGVKRRNMDDM-AGRPMVKGGSGEDKVPTQQKRHQLNNFGQM      174
                   :: :  :     : :      :::  :
Db   3635  AESVKDEADKSKEESRRESGAEKSPLASMEASRPTSVAESVKDETEKSKEESRRESVTEK  3694

Qy    175  RQTMLSKMAHPASA--NAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVEIAQE   232
            :::  :    :      ::: :  :   :::   ::
Db   3695  SPLPSKEASRPTSVAESVKDEAEK-------SKEESRRESVAEKSPLASKESSRPASVAE  3747

Qy    233  DDDSEFQQLHQQRLARERENPQQPPKL---GVATPISARFQPKLTAVAESVLEGTDTTQ    288
              : :                          : : :      : ::::
Db   3748  SIKDEAEGTKQE--SRRESMPESGKAESIKGDQSSLASKETSRPDSVVESVKDETE---  3801

Qy    289  SPLKPQ-SMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHAS   347
               : :    : :  :      :  :
Db   3802  ---KPEGSAIDKSQVASRPESVAVS------AKDEKSPLHSRPESVADKSPD---AS    3846

Qy    348  SDGSQHLLLDNKGHLFDIKSTATS------YSVLHNSHP----GEIKGKLAQAGTGSVSV   397
              :   :        ::          :  : :       : :  :
Db   3847  KEASRSL--------SVAETASSPIEEGPRSIADLSLPLNLTGEAKGKLP---TLSSPI   3894

Qy    398  DGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHIL   457
            :   ::   :           ::::       : :  : ::
Db   3895  DVAEGDFLEVKAESSPRPAVLSKPAEFSQPDTGHTASTPVDEASPVLEEIEVVEQ-----  3949

Qy    458  HPELGVWQSADKDTHSQLSRQADGKLYALKDNR--TLQNLSDNKSSEKLVDKIKSVSVDQ   515
            :   ::                :::::                  : :
Db   3950  HTTSGV--------GATGATAETDLDLTETKSETVTKQSETTLFETLTSKVES-------  3995

Qy    516  RGQVAILTDTPGRHKMSIMPSLDASPESHISLSHFA-----DAHQGLLHGKSE--LEAQ   568
            : : :      :  :           ::
Db   3996  ------------KVEVLESSVKQVEEKVQTSVKQAETTVTDSLEQLTKKSSEQLTEIK   4041

Qy    569  SVAISHGRLVVADSEGKLFSAAI-----PKQGDGNELKMKAMPQHALDEHFGHDHQISGFF  624
             :  :  :   :      :             ::   :
Db   4042  SVLDTN----ISNVTNLFSTAVETIEKKVQDVTEKVIEKATEH-VSEHVTTGESSTET   4095

Qy    625  HDDHGQLNALVKNNFRQQHACPLGN                                     661
             :::                     :::
Db   4096  SQEKSSLDLGTFSELRETHITTVGSPEFTVTICERDEPVLHDIKEDEEHRFSPPSDVDK  4155

Qy    662  ALVIDNQLGLHHTNPEPHEILDMGHGLGSLALQEGK------LHYFDQLTKGWTGAESDCKQ  716
            :  :        :        : :: :               : : :  ::
Db   4156  AAIIPPQ-PMRPLSPREEEVAKIVADVAKVLKSDKDITDIIPDFDE------------RQ  4202

Qy    717  LKRGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDK   776
            :   :         ::  :     :       :  ::   :  : :
Db   4203  LEEKLKSTADTEEESDKSTRDEKSLEISVKVEIESEKSSPDQKSGPISIEEKDKIEQSEK  4262

Qy    777  AQAMAVIGVNKYLALTEKGDIRSFQIKPGTQOLERPAQTLSREGISGELKDIHVDHKQNL   836
            :  : :                :        :           :
Db   4263  A----------------QLRQGILASSRP----ESVASQPESVPSPSQS-           4291

Qy    837  YALTHEGEVFHQPREAWQNGAESSSWHKLALPQS-ESKLKSLDMSHEHKPIATFEDGSQH   895
            :  : :        :    ::        ::    :    :      :
Db   4292  -AASHE---HKEVEL----SESHKAEKSSRPESVASQVSEKDMKTSRPASST---SQF   4338

Qy    896  QLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGA   955
            :          :                       :    :
Db   4339  STKEGDEETTES----LLHSLTTTETVETKQME-------------EKSSFESVSTSVT  4380

Qy    956  EGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLD  1015
            :::: :    :     :   :         ::
Db   4381  KSTVLSSQSTVQLREESTSESLSSSLKVEDSS------RRESLSSLLAEKGGI-        4427

Qy   1016  AHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLQOHQGVLKS  1075
             ::                       :::
Db   4428  ATNTSLKEDTSASASQLE-------ELL-----VQSEECSSESIVSEIQTSIAQKS     4471

Qy   1076  NGEI------NSEFKPSPGKALVQ-                                    1103
             :::            :
Db   4472  NKEIKDARETKVTSQFTTTSSATKDDSLKETVAEFLATEKIVSAKEAFSTEATKSADDC  4531

Qy   1104  LSKSLQQAVHATPPSA----ESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDK     1155
            :    :             :::: :
Db   4532  LKKTTASAVSSTSASQRALFVGTDESRRESLLSQASESRLTHSDPEDEPAD---DVDER  4588

Qy   1156  TALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKH  1215
             : :                        :
Db   4589  SSVKESR----SKSIATI-----MMTSIYKPSEDM-                        4617

Qy   1216  YTDMGFTHNKALEANYDAVKAFINAFFKEHHGVNLTRTVLES--QGSAELAKKLKN-     1270
             : :                         :  :
Db   4618  ------------SKLVEEEHEVE---ELAQEVTSTSKTT-TLLQSSEQSSTTTSSTSKTGAS  4664

Qy   1271  ----TLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSR  1325
            :  :                  :
Db   4665  RVESITLTQMD----QQTSQSGPDADRKTPPT----APVSPGV-                4700

Qy   1326  TSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAA  1385
             :::
Db   4701  --------------KAMSSTG--------------SAGSV----                4716

Qy   1386  VSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIEHQMKGGSKLTFSVDT  1445
```

```
Db  977  QAKIINRGTINV------KNKLEYGSNVDV-----ENNMRSMQVNL---YEKIFNGDNPI---  1022
          :|          |                       |::                         |
Qy  569  SVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHF---GHDHQISGF        623
          ::|:|                               :|                |:||
Db 1023  TLTLKNGVTFAKDFSNRRRRAS-----NDGEGTNKKTFDNVAHLIEEAFSGYSNGNDHRAS-    1078
              ||                              ::|                    |::
Qy  624  FHDDHGQLN----ALVKNNFPQQHACPLGNDHQFHPGWN-LT---DALVID-NQ            668
          |                  |:|    ::  :|                         ::
Db 1079  ---DDGHVKSPYYLLVLAQAVNNTEGENVLKTALQHIFGPGNWNDLTTTNNDTTTINDKWNQ   1135
              |                    |:|                |::
Qy  669  LGL-------HHTNPEPH--EILDMGHLGSLALQEG---------------------KL       697
          |          |                        |                    |
Db 1136  LKLKWEKFKNNGENNHSINLNIYPADEGVEKAKIFAGVLRNGTNGVEDKVYQELNDKAKK     1195
Qy  698  HYFDQLTKGWTGAESDCKQLKKG-----------LDGAAYLLKDGE---------VKRLNINQS  741
          :|                      |                        :|
Db 1196  EYEDKFAKKFQGRFK-----SRFQNGEFDWAGDWAKEGNESYGSKETEEKYNGIKKEHTVNIG  1253
          ::|                     |                       |::
Qy  742  TSSIKHGT--ENVFSLPHVRNKPEPGDALQGLNK-----------------------         773
           |  |        |       :  :||
Db 1254  KHEIKVPTVSFPENLNNINHQQDKSD----------GIDKSIISELLAQPIYVAKADVPDVDPRV 1307
Qy  774  --DDKA-----QAMAVIGVNKYLAL-----------TEKGDIRSFQIKPGTQQLERPAQT       815
            ||:                                |||
Db 1308  AQNDKAVDEDGLYRTRLSYINQNNYLGAKYFFNQLDTEDDKLKGIK-RIGDNYFEH--QL     1364
          :|:||                                ||:           :|
Qy  816  LSREGISGELKDIHVDHKQNLY--ALTHE-----------GEVFHQPREAWQNGA          857
           |    |||                                 |
Db 1365  ITR--LIEKVADNHLTLKHGLHDIALVKKLIDSASIQAKDLNLKVGEAL--TKEQKDNLK    1420
Qy  858  ESSSWH----------KLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPE   909
          :                 |                                |
Db 1421  EDIVWYVKTEVNAQEVLVPQVYLAKQTIEEVEKQRGVT---------GQIRAGIIDVKVDDV  1474
Qy  910  RGP---------LAVGTSGSQTVFN-----RLMQGVKGVKIPGSGLTVKLSAQTGGMTGAEG   957
          |:|           |||                   |::|
Db 1475  RNTGTIAGYAVGLEAKNKLKNTGDILSQRLSKLVGKKGLESTGVT---YVDETG--ATKV    1529
Qy  958  RKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAH    1017
          ||                                   :|::                  |
Db 1530  RKARIKSEGHI-----------YLETDKD-KNVD-LTASELKGNTG-----------QIKAK  1567
Qy 1018  NVRHN-APQPDLQSKLETLDLGEHGAELLNDMKRFR--------DELEQSATRSVTV        1065
          ::  :         |:   |   ::|                    :|
Db 1568  DLNLNDIYETSYKYKYEKL-FGKNGGEIGDRVTQTSQAKSVGTDASFDHLHLSLEGDVNQ    1626
Qy 1066  LGQHQGVLKSNGEINSEFKPSPGKAL-----------VQSFNVNRSGQDLSKSL-QQAV     1112
          ||  |                                 |:
Db 1627  TGSNLKANRTTGVVKGDFNTKAGKDLFHRQIDTVTSGTVYSASASGGGQSAGISLTDQGV    1686
Qy 1113  HA-TPPSAESKLQSMLGHFV-------SAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLLDT   1167
           |  |                         |:|
Db 1687  ETYTNKTATAGANADVTNFMKRTRETETSLTHRNSEF-----------NALSGELYVMGK    1735
Qy 1168  VTIGEL-----------HELADKAKLV--------------SDHKPDA               1190
          |:  |             |:  ||
Db 1736  ADIGGVDINRDVEVIKTPEIAAEQKAAEEAKKAEVKENEASETAAKETEEAENDNVAEK     1795
Qy 1191  DQIK-----QLRQQFDTLREKRYESNPVKHYTD--MGFT-------HNKA--LEA        1229
          |:|        |                        ||            :|
Db 1796  DKTKPKFKKLTDEIAAAFETKGEDFFAAYKAREEDRKKGFTLSAEQIESTKARDEKET      1855
Qy 1230  NYDAVKAFINAFKKEH--------HGVNLTRTVLESQGS-----------A             1262
          ::| |                    ||                        |
Db 1856  TYYELKVGVGAEAEAHSAAADAISNKARQIIDTQNGLKQDGTVALQEASDVLNLATGDLA    1915
Qy 1263  ELAKKLKNTLLSLD-SGESMSFSRSYSYGGG----------VSTVFVPTLSKKVPVP       1306
          ||   |          |                         |:
Db 1916  GASAKLKFELSTIEKKSRGASDGRSILGGRLNLAARGDITLNNVETTENSHLSLKARDN     1975
Qy 1307  VIPGAGIT--LDRAYNLSFSRTSG---GLNV------SFGRDGGYSGNIMVATGHDVMPYM    1356
          |:                                     |
Db 1976  VNVNSGVTEQKDESNSQSLKVTAGASSGCGVMAGGCSAGVSAGVSGS----------YN     2024
```

```
Db 4717  AGAVAAGGKCESSAASIVS----------:|:||:-SSGPMSP------------KDISGKSSPGALTSESQSI  4761
          :|:||:                         |                      |
Qy 1446  SANLDRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTSASNNR      1505
Db 4762  PTPLGRESHTDTPESSPKPTSPFPRVSKDELKSLEMQHHSQEQMLAGAAAAGAECEGDI      4821
          |::                                :|
Qy 1506  PTF--LNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNR---TSQSI    1560
          |:                                                       :|
Db 4822  PELHELRGL---ECTTALS---GSTDK--------IITTTITTVTKVISADGKEIVTEQKT    4868
Qy 1561  SLELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAALKEL--------DD-----AKP     1606
          ||:|                                      :|          |
Db 4869  VTTTDSSEPDSEKVVVTTTRTTSESERDQLLPKEVALLRGLYRASTPGSEDDEDLLLGSP     4928
Qy 1607  AEQLHILQQHFSA--------KDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTT    1659
          :|                   |:|:                                :||
Db 4929  RSATSYELQHSSSGVSKRSDLDADGDE----------SQDDIPPQYGSEEHSTA          4972
Qy 1660  YNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMNNNDPALKDIIKQLQSTPF----     1713
          :|:                                              |  |
Db 4973  RSILLPRTADPMA----TSFYGALPDSF----------DVVMKPSTEPIPIQGAP        5013
Qy 1714  ---SSASVSMELKDGLREQT-------EKAILDGKVGREEVGVLFQDRNNLRVKSVSVSQ   1763
              ||                     |||
Db 5014  SGDSQSSSESVESS-----SQTWAGHKFLDQADKDFQRALEE-----HVQARGAEVMSSVTAKY 5066
Qy 1764  SVSKSEGFNTPALLGTSN----------SAAMSMERNIGTINFKYGQDQNTPRRFTLEGG   1814
          |:|                          |||
Db 5067  SYSPSKAEEMEQIVSGTAERQRFPLSDVQRARVAESGFATVGSVASQQQQQEK---GGE     5122
Qy 1815  IAQANP-----------------QVASALTDLKKEGLE                         1835
          :|:||                   |
Db 5123  VEQAVPTTTAVTASTTATASSTGALPKDRLE                                 5153
```

RESULT 11
T31102
filamentous hemagglutinin 1 - Haemophilus ducreyi
C;Species: Haemophilus ducreyi
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 22-Oct-1999
C;Accession: T31102
R;Ward, C.K.; Lumbley, S.R.; Latimer, J.L.; Cope, L.D.; Hansen, E.J.
J. Bacteriol. 180, 6013-6022, 1998
A;Title: Haemophilus ducreyi secretes a filamentous hemagglutinin-like protein.
A;Reference number: Z20984; MUID:99030326
A;Accession: T31102
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-4152 <WAR>
A;Cross-references: EMBL:AF057695; NID:g3929017; PID:g3929018; PIDN:AAC79757.1
C;Genetics:
A;Gene: lspA1

```
Query Match         2.5%;    Score 237.5;  DB 2;   Length 4152;
Best Local Similarity 18.9%;  Pred. No. 0.0052;
Matches  345;  Conservative 244;  Mismatches 636;  Indels 599;  Gaps 88;

Qy  350  GSQHLLLDNKGHLFDIKSTATSYSVLHNSHP------------GEIKGKLAQAGTGSVSV    397
          | |||                                       |:|
Db  835  GGQGLNLTAKGNITN-DSNSTAIAVLHSNNDINLNANNKVYNIGEIIYSQ----AGNISV    888
Qy  398  DGK-----SGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRL      449
          |||      |:|                                          ||:|
Db  889  EAKLLHNDVKLSGNIT-TTTKSGNATVKT--------NSIGGGLHDANSIRV           931
Qy  450  HDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIK    509
          |                 |:|   :|                                |
Db  932  GELTLN------GKFADLD--NQLKVALRGKIYA-GSNLIFK----AKEGEKEQKSTA     976
Qy  510  SYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHI-SLSLHFADAHQGLLHGKSELEAQ   568
```

```
Db   335 NQIKASDLMGDDITLQGADLTIDGKQLQOKETDIDNRWFYSWKYDVTKEKEQIQQIGSQI  394
Qy   480 DGK----LYALKDRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMP   535
         :     :     ::: ::     ::    :  ::      :          :.:
Db   395 DAKNNATLTATKGDVTLDAAKINAGNNLAINANKDIHING-----LVEKESR----     441

Qy   536 SLDASPESHISL--SLHFDAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPK   593
         :|  |          || |:: ::|        | :.      ::     :|: :::|:|
Db   442 SENGNKRNHTSRLESGSWSNSHQETLKASELTA------GKDLGLDAQGSITAQGAKL    494

Qy   594 QGDGNELKMKAMPQHALD-EHFGHDHQISGFFHDDH----GQLNALVKNNFRQQ---   642
         : | ||  :   |:         :|    :           ::||
Db   495 HANENVL-VNAKDNINLNVQKTNNDKTVT---DNHVMWGGIGGQN--KNNNNQQQVS    546

Qy   643 HACPLGNDHQFHPCWNLTDALVIDNQLGLHHTNPEPHEILLDMGHLGSLALQEGKLHYFDQ  702
         || | |||  |:    |||  :   |                                :
Db   547 HATQLTADQGL------LLAADNNVNI---                               567

Qy   703 LTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLN-INQSTSSIKHGTENVFSLPHVRNK   761
         ||:   |:        :|   |  : |    :  .   : : :: :: ::   :
Db   568 ---TGSQ----VKGNQGAFVKTTQGDVVIDNALSETISKIDERTGTAFNITKSSHK      616

Qy   762 PEPG-DALQGLNKDDKQAQAMAVIG--VNKYLALTEKGD------IRSFQIKPGTQ--QLE  810
         .: | :| :|   |.   :   :|:   |||  :::  |      :|:|: | ::
Db   617 NETNKQTSTGSELISDAQLTVVSGNDVNVIGSLIKSADKLGIHSLGDINVKSAQQVTKID   676

Qy   811 RPAQTLSREGISGELKD------IHVDHKQNLYALTHEGEVFHQPREAWQNGAESSSWHK   864
         : |: ||  |:|  ||:        |:||| |  || ||                : ::
Db   677 DEKTSLAITGHAKEVEDKQYSAGFHITHTTNKNTSTE---                     713

Qy   865 LALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTVF   924
         :  :.|:.   |:: |:| : :|  :||||                         |
Db   714 -TEQANSTISGANVDLQANKDVTFA-GSDLKTTAGN----ASITGDNVAF          757

Qy   925 ---NRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTP   980
            | |: ||.||:| |:||| ::|:: |||
Db   758 VSTENKKQTDNTDTTISGG------FSYTGGVD-----KVGSKAD----FQYDKQHTQT   801

Qy   981 RPIKNAAYATQ------HGWQGREGLKPLYEMQGALIKQLDAHNVRH             1021
         ::| |   |::        |:||||                     :
Db   802 EVTKNRGSQTEVAGDLTITANKDLLHEGASHVEGR---YQESG----ENIQH        847

Qy  1022 NAPQPDLQSKLETLDLG-EHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEIN  1080
         .| :|  ::::|||| |  :.| :|::                     ||  :|:|:
Db   848 LAVNDSETSKTDSLNVGIDVGVNL--------DYSGVTK------                878

Qy  1081 SEFKPSPGKALVQSFNVNRSGQ-DLSKSLQQAVHATPPSAESKLQSMLG-HFVSAGVDM  1137
         .  | : || :    |::|  :  ||::| ::                : : |:|| :
Db   879 --PVKKAIEDGVNTTKPGNNTDLIKKV---TARDAIANLANLSNLETPNVGVEV       927

Qy  1138 SHQKGEIPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLR  1197
         |::||:|  |: ::|:||||||:|||:||||:|||:|:|||||||:||||| ||:|:|||
Db   928 GIKGG---GSQSQTDSQAVSTS----INAGKIDIDSNNKLH                   962

Qy  1198 QQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTRTVLE  1257
         ::     :  :     ::  ::| :|:|:||     ||    :::::  |||:| ||:
Db   963 DQ------GTH------YQSTQEGISLTANT---                          981

Qy  1258 SQGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDR  1317
         |.||| |||:| |||||:||||:||||||||||||| : |: ::: |   |||  :||
Db   982 ---------HTSEATLDKHQI-TFHETKGGG--QIGVSTKTGSDITVAIKGEGQTTDN   1027

Qy  1318 AYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKIS  1377
         |  : ::: :|| ||  |  :   | | :: |:    |: |:  :|||:||  :||
Db  1028 A--LMETKAKGS---QFTSNGDISINVGENAHYEGAQFDAQKGKTVINAGG-----    1073

Qy  1378 PDLRIGAAVS--GTLQGTLQNSLKFKL----TEDELPGFIHGLTH------GTLT     1420
         |||||||| |  | ||   ::|::      ||::    |:  : |       |||:
Db  1074 -DLTLAQATDTHSEQSQNVNGSANLKVGTTPESKDYGGGFNAGTTHHSKEQTTAKVGTIT  1132

Qy  1421 PAELLQKGIEHQMK-QGSKLTFSVDTSAN----LDLRAGINLNEDGSKPNGVTARVSAGL  1475
         .|:|:::|:|:|:|  |||:|||||:|:|    |||||||||||||||||||:| |||||
Db  1133 GSQGIELNAGHNLTLQGTHLSSEQDIALNATNKVDLQSA--SSEHTEKGNNLSGGVQAGF  1190
```

```
Qy  1357 TGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTH  1416
         :|::|: ::     ::  | :| :  || :|                 |::|||:  ||
Db  2025 ESNTESTSHTNSLLRGK----GKDFNLISSNVDVDHLHLDVKGDTN               2072

Qy  1417 GLTPAELLQK--GIEHQMKQGSKLTFSVDTSANLDLRAGINLNEDGSK---PNGVTAR   1470
         |: :: :|::  |:    :||   ||||:| |  ::::| :   :::|   |||||:
Db  2073 VVSKQDSYSRKERGVNYSVSAGVGVSTAGGARPNGSVGLGVSAENENSKIVKQQAGISAK  2132

Qy  1471 VSAGLSASANLAAG------SRERSTTSGQFGSTTSASNNRPTF                1508
         :|||  ||:|| :         |||| ||   |:  ::  ||
Db  2133 RITGEINNLNLTGGYIENKGNPDELNVKGDITTHELKDEHHKDGGSFGGSVGVSETGVTQ  2192

Qy  1509 LNGVGAGNLTRAALGVAHSS----THEGKPVGIFPAFTSTNVSAALALDNRTSQSISLEL  1564
         :|  |  ::  ::  || ||    |:|||  |:   :|   ::     ||||||: :: |
Db  2193 VNVNGGRVEQKHYEATQHSSISGINTKGKTVGNF------KTDRSQSTEV          2236

Qy  1565 KRAEPVTS-------NDISELTSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHF  1617
         ||::    :       ||: | :: ||:||:|||      :|:||::||
Db  2237 HRDTIAATNFNFELGDIAELAKK-GKEKWDNRSAKTTS--------               2274

Qy  1618 SAKDVVGDERYEAVRNLKKLVIRQQAA------DSHSMELGSASHSTTYNNLSRINNDGI  1671
         ::|| |  :  : ||| :: |:|||       ||:|||||||:|:||| |||:||:||:|
Db  2275 SSQDSAHDPRSRSVENGYSELPRFKTADNDAGVDSPRLIKGEAEQAQTL-ALTKAGNDVI  2333

Qy  1672 VELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTE  1731
         :|||::|:||| |:|||||||||||||||||||||||:|||:|:|||:||:|||||:|||
Db  2334 -------------PEVQSLTQKARPQSLVDESPYAEIPALTRPQVK              2366

Qy  1732 KAILDGKVGREEVGVLFQDRNNLRVKSVSVSQ--SVSKSEGFNTPALLLGTSNSAAMSM   1788
         .::: |:::::::  :|:|::||:|:|||||    |: : :||  :::  ||:|||| |
Db  2367 SNI-------AESIEVPQFRTKVSDGDEGNYAEITFPT-NKAAISS                2404

Qy  1789 ERNIGTINFKYGQDQNTPRRFTLE                                    1812
         ::|:|| :   ||   |: ||:|:
Db  2405 TQDVG-------DTPTPRALRLE                                       2420
```

RESULT 12
A35140
hemolysin A precursor - Proteus mirabilis
C;Species: Proteus mirabilis
C;Date: 03-Aug-1990 #sequence_revision 03-Aug-1990 #text_change 08-Oct-1999
C;Accession: A35140
R;Uphoff, T.S.; Welch, R.A.
J. Bacteriol. 172, 1206-1216, 1990
A;Title: Nucleotide sequencing of the Proteus mirabilis calcium-independent hemolysin ge
A;Reference number: A35140; MUID:90170827
A;Accession: A35140
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1577 <UPH>
A;Cross-references: GB:M30186; NID:g150888; PIDN:AAA25657.1; PID:g150890

```
Query Match        2.5%;  Score 232;  DB 2;  Length 1577;
Best Local Similarity 18.9%;  Pred. No. 0.002;
Matches  329;  Conservative 238;  Mismatches 614;  Indels 564;  Gaps 84;

Qy   261 VATPISARFQPKLTAVAESVLEGTDTTQ-----SPLKPQSMLKGSGAGVTPLAVTLDKG   314
         :|  | :::| | :||| : : ||:||      |.| ||||| ||: ||||||| :|
Db   197 IAPRIDSR--GKITAAEISAFTGQNTFSQHFDILSSQKPVSALDSYFFG-----SMQSG   248

Qy   315 KLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSV  374
         :: :    :   : ||:| || : :  : .    | || ::  ::                :
Db   249 RIRI---INT-----AEGSGVKLAGKFADNDLSVKADN------IQTDSQV         286

Qy   375 LHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQ  434
         :::|   ::.| ||| ||||||  ||||  || || |::|: |  | :   :::: |
Db   287 RYDSYDKD--------GSENYQNYRGGITVNNSGSSQTLITKTELKGK--NITLVASSH   334

Qy   435 HPAGAARPQGESIRLHDDKIHILHPELG-----------VWQ----SADKDTHSQLSRQA  479
         | |:|||||||| || :| ::||| |            :||:    |:||||:||||: |
```

```
Qy    70 AHQQKKSFSLRGCLGTKKFSRSAP-----------------------------OGQP-----GTT 100
            :|        :|              |||:||||||                :|:
Db   117 NRFKQFDVDEKGVILNNSRSNTQTOLGGWIQGNPHLARGEARVIVNQIDSSNPSLLNGYI 176

Qy   101 HSKGATLRDLLARDDGETOHEAAAPDAARLTRSGGVKRRNMDDMAG------RPMVKGGSG 155
         ||     :|:      |||        :|   ::|
Db   177 EVGGKRAEVVVANPSGIRVNGGGLINAASVTLTSGVPVLNNGNLTGFDVSSGKVVIGGKG 236

Qy   156 EDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAG----------DRLQHSPPHIPG 203
            |       ::|     |   |    |:||                |||
Db   237 LDTSDADYTR----------ILSR-------AAEINAGVWGKDVKVVSGKNKLD----FDG 276

Qy   204 SHHEIKEEPVGSTKATTA-------------HADRVEIAQEDDDSEFQQLHQQRLARERENP 253
            ::|          |||     :|:| :|: |      |||     |
Db   277 SLAKTASAPSSSDSVPTVAIDTATLGGMYADKITLISTDNGAVIRNKGRIFAATGGVTL 336

Qy   254 PQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLK--PQSMLKGSG---------AGVTP 306
         |   ::|    ||                  :|:|          ||:
Db   337 SADGKLSNSGSIDA-------AEITISAQTV----DNRQGFIRSGKGSVLKVSDGINNQAGLIG 389

Qy   307 LAVTL---DKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLF 363
            ||         :|  :||         :|| |:              |||:
Db   390 SAGLLDIRDTGKSSLHINNTD--GTII----AGKDV-----------SLQAKSLDNDGILT 433

Qy   364 DIKSTATSYSVLHN-----------------SHPGEIKG-KLAQAG----------TGSVSVDGKS 401
            || :|   :|               :||  |                  |||
Db   434 AARDVSVS---LHDDFAGKRDIEAGRTLTFSTQGRLKNTRIIQAGDTVSLTAAQIDNTVS 490

Qy   402 GKISLGSGTQSHNKTMLSQPGEAHRS-----LLTGIWQHPAGAARPQGESIRLHDDKIHIL 457
         |||    |       |    ||     :| :|:         || |||  |||
Db   491 GKIQSGNRTGLNGKNGITNRGLINSNGITLLQTEAKSDNAGTGRIYGSRVAVEADT---LL 548

Qy   458 HPE-----------LGVWQSADKDTHSQLSRQA----------DGKLY-ALKDNRTLQNLSDN 498
         ||         :|              :|:|         |    |  :|
Db   549 NREETVNGETKAAVIAARERLDIGAREIENREAALLSSSGDLHIGSALNGSRQVQGANTS 608

Qy   499 KSSEKLVDKIKSYSVDQRGQVAILT-----------------DTPGRHKMSI- 533
         |||              |||     :|      :|: |||:||||
Db   609 LHN-------RSAAISSGNIRIATKDLQNTNEHLRFHTEETHREHRIEYQAEGRTERYP 661

Qy   534 ----------MPSLDASPES-----------HISLSLHFADAHQG-LLHGKSEL 565
                   |||:||||         :|  :|||:||  |||
Db   662 EGSQKELGWEIFEDESLHMRTPDGSPHSVWYKDYERITAESKITESKPGQIISGGNLVL 721

Qy   566 EA-----------QSVAISHGRLVVA----------DSEGKLFSAAIPKQGDNELKMKAMPQHALDEH 613
          :|    :| ::||:        |    :|| |||:||
Db   722 DAAKLKNHNSRIIAGGRLIVGTPESALDNDETLGTKTITDKGD-----------LHRYRH 771

Qy   614 FGHDHQISGFFHDDHG----------QLNALVKNNFRQQHA----------CPLGNDHQFHPGWN 658
         |   :|:|||          :|  ||     |||  :||         |:|||  :|||||
Db   772 HKKGRDSTGYSRSPYEPAPEVSSIRMGISAYKGYAPQQASDIPGTVPVVAENGIHPFFT 831

Qy   659 LTD-----------ALVIDNQLGLHHTNP--EPHEILDMGHLGSLALQEGKLHYFDQLTKGW 707
         |:|          ::|  |||:|||:  |||  ||||:|:|||:|||  :|||||| :|||
Db   832 LPNSSLFAIAPNNKGYLIETDPAFTDYRKWLGSGYM--LAALQODPNHIHKRLGDGYYEQK 890

Qy   708 ------------------TGAESDCKQLKKGLDGAAYLLKD----------GEVKRLN 737
                       |||    :| :|||:|||  :| ::| |||
Db   891 LVNEQIAKLTGYRRLDGYTNDEEQFKALMDNGITIAKELQLTPGIALSAEQVARLTSDIV 950

Qy   738 -INQSTSSIKHGTENVFSLP--HVRNKPE----------PGDALQGLNKDKAQAM---------- 780
          :|  ||    :|  |||:   :| |||        ||  ||:| |||:|||
Db   951 WLENETVTLPDGTTQTVLKPKVYVRARPKDMNGQGALLSGSVVDIGSGAIENRGGLIAGR 1010

Qy   781 -AVI----------GVNKYLA--LTEKGDI-RSFQIKPGTQQLERPAQTLSRE 819
          :|      :|  |:|     ||  :|  |  :|:|   |||     :|
Db  1011 EALILNAQNIKNLQGDLQGKNIFAAAGSDITNTGSIGAENALLLKASNNIESRSETRSNQ 1070

Qy   820 GISGELKD-----------IHVDHKQNLYALTHEGEVFHQPREAWQNGAESSSWHKLALPQSESK 873
          :||:|:        :|:::| |||:|  |:|  :|           ::|::| |||
Db  1071 NEQGSVRNIGRVAGIYLTGRQNGSVLLDAG----------NNIVLTASELT 1111
```

```
Qy  1476 S----------------ASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAAL 1522
            :|     :|     |||:||||      :|:
Db  1191 GKKMTDDASSVNGLGSAQFAIGKQDEKSVSREGGTINNSGN-----------------L 1232

Qy  1523 GVAHSSTH-EGKPVGIFPAFTSTNVSAALALDNR-TSQSISLELKRAEPVTSNDISELTS 1580
         |:   :||        :| :|:|||      ::| || |
Db  1233 TINGNSVHLQGAQVN----------SKDTQLTSQSGDIEITSAQ--STDYKNNWG 1275

Qy  1581 T-LGKHFKDSATTKMLAALKELDDAKPAEQLHIL----------QQHFSAKDVVGDERYE 1629
         | :|:::|          :|||:|| |                 ::|:||
Db  1276 TDIGFNGKKTNNTP------KEVTEEKPATSIHNIGGKLLVVEDQQKTSHQNATLETGTL 1330

Qy  1630 AVRNLKKLVI--RQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPASS 1687
         |: ::|     :|:         ||:         |     |:|:||   :|    :|
Db  1331 TINSNKDLTLSGANVTADSVTGNVGGSLNIASQKESDRHVTVG-VNVGYNHTND--PKSS 1387

Qy  1688 ---AKRLGEMMNNDPALKDII----------KQLQSTPFSSASVSMELKDGLREQTEKAIL 1735
            |::|   :|:|  :|:|        :|  |:| :::|:| ||   |:|| :| |||
Db  1388 QVNKTAKAGGSLL--EKTIKDTIDSGIKSSTDAISDKYNSLSSTIADKTGISDET-KAKI 1444

Qy  1736 D---GKVG------------REVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALL 1777
         |   ||:|               :|  |:| |::| ::  |:::| |:|     |||
Db  1445 DQGFGKVGNGIKNIVTGAEGHTANADIKVTHVD-NDAVTKTTSLTSNNDLSLNVNGSTKL 1503

Qy  1778 LGTS------------NSAAMSMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASALT 1827
         ||:              :|:       |:            |||     :|
Db  1504 TGAEIVSQQGQVDLGGSSVKLE-NIEGHHYEAGADLD-----------------LKSSVV 1545

Qy  1828 DLKKE 1832
         ||: :
Db  1546 DLAKQ 1550
```

RESULT   13

T09083

hemagglutinin/hemolysin-related protein NMB1214 [imported] - Neisseria meningitidis (gro

N:Alternate names: probable secreted protein

C:Species: Neisseria meningitidis

C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2000

A:Accession: T09083; B81109

R:Simpson, N.J.; Spratt, B.G.

submitted to the EMBL Data Library, October 1997

R:Reference number: Z16558

A:Accession: T09083

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-2273 <SIM>

A:Cross-references: EMBL:AF030941; NID:g2623257; PID:g2623258

A:Experimental source: strain 44/76

R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;

ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755

A:Accession: B81109

A:Molecule type: DNA

A:Residues: 1-2273 <TET>

A:Cross-references: GB:AE002469; GB:AE002098; NID:g7226446; PIDN:AAF41596.1; PID:g722645

A:Experimental source: serogroup B, strain MMD58

C:Genetics:

A:Gene: pspA; NMB1214

```
Query Match         2.4%;    Score 229.5;   DB 2;    Length 2273;
Best Local Similarity 18.9%;    Pred. No. 0.0048;
Matches 419;   Conservative 285;   Mismatches 790;   Indels 725;   Gaps 107;

Qy   13 AVHTAAHNPV-GHGVALQQGSSSSSPQNAAASL--AAEGKNRGKMPRIHQPSTAADGISA 69
          :|::   |: |:|:|                  :|  |:|::           ::|
Db   66 ALGVAAFSPAPASGII----------ADKSAPKNQQAVILQTANG----LPQVNIQTPSSQGVSV 116
```

```
RESULT 14
A28182
hemolysin A - Serratia marcescens
C;Species: Serratia marcescens
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Oct-1999
C;Accession: A28182
R;Poole, K.; Schiebel, E.; Braun, V.
J. Bacteriol. 170, 3177-3188, 1988
A;Title: Molecular characterization of the hemolysin determinant of Serratia marcescens
A;Reference number: A28182; MUID:88257037
A;Accession: A28182
A;Molecule type: DNA
A;Residues: 1-1608 <POO>
A;Cross-references: GB:M22618; NID:g340726; PIDN:AA50323.1; PID:g556420

Query Match          2.3%;   Score 221;   DB 2;   Length 1608;
Best Local Similarity  18.6%;   Pred. No. 0.0072;
Matches   372;   Conservative 235;   Mismatches 689;   Indels 700;   Gaps 89;

Qy   33  SSSSPQNAAASLAAEGKNRGKMPRIHQPSTAA-------DGISAAHQQKKSFSLR--GC  82
Db   23  AASAGAYAAEIVAANGANG---PGVSTAATGAQVDIVAPNGNGLSHNQYQDFNVNQPGA  79

Qy   83  L--GTKKFSRSAPQGQPGTTHSKG-----ATLRDLLARDDG--------ETQHEAAA 124
Db   80  VLNNSREAGLSQLAGQLGANPNLGGREASVILNEVIGNPSLLHGQQEIFGMAADYVLAN 139

Qy  125  PDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKRHQLNNFGGMRQTMLSKMAH 184
Db  140  PNGISCQSCGFINTSHSSLVVGNPLVENGVLQG-------YSTFGNRNTLSLN------ 185

Qy  185  PASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRVE-IAQEDDDSEFQQLHQ 243
Db  186  -GTLNAGGVLDLIIAPKIDSRGEVIVQDFKQSNGKVTSAAINAISGLNRVARDGTVQASQQ 244

Qy  244  QRLARERE-----------------NPPQPPKLGVATPISARFQPKLTAV--------AESV 280
Db  245  MPTALDSYYLGSMQAGRININTAQGSGVKLAGSLNAGDELKVKAYDIRSESRVDDASSN 304

Qy  281  LEGTDTTQ-------------------SPLKPQSMLKG----------SGAGVTPLAVTLDKG 314
Db  305  KNGGDNYQNYRGGIYVNDRSSSQTLTRTELKGKNISLVADNHAHLTATDIRGEDITLQGG 364

Qy  315  KIQLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSV 374
Db  365  KLTL-------------------DGQQ--LKQTQGHTDDRWFYSWQYDV 392

Qy  375  LHNSHPGEIKGLKLAQAGTGSVSVDGKSGKIS--------LGSGTQSHNKTMLSQPGEAH- 425
Db  393  TRE-------REQLQQAGS-TVAASGSAKLISTQEDVKLLGANVSADRALSVKAARDVHLA 445

Qy  426  ----------------------PAGAARPQGE----------SIRLHDDK 453
Db  446  GIVEKDKSSERGYQRNHTSSLRTGRWSNSDESESLKASELRSEGELTLKAGRNVSTQGAK 505

Qy  454  IHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQ--NLSDNKSSEKLVDKIKS 510
Db  506  VHAQRD-----LTIDADNQIQVGVQKTANAKAVRDDKTSWGGIGGGDNKNNS------- 552

Qy  511  YSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQ- 568
Db  553  NRREISHASELTSGGTLRL-----------NGQQGVIITGSKARGQKG 589

Qy  569  -SVAISHGRLVVADSEGKLFSAAIPKQGDGNELRKMKAMPQHALDEHFGHDHQISGFFHDD 627
Db  590  GEVTATHGGLRIDNA------LSTTVDK---------IDARTGTAFNITSSSHK- 628

Qy  628  HGQLNALVKNNFRQQHACPLGNDHQFHPGWNLT-----DALVIDNQLGLHHTNPEPHEIL 682
Db  629  ------ADNSYQSSTASELKSD-----TNLTLVSHKDADVIGSQVA-------- 663

Qy  683  DMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNIQST 742
```

```
Qy   874  LKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSG----SQTVFN-----  925
Db  1112  NQS------------------EDG-QTVLNAG-------GDIRSDTTGISRNQNTIFDSDNY 1147

Qy   926  RLMQGVKGKVIPGSGLITVKLSAQ------TGGMTGAEGRKVSSKFSERIRAYAFNPTMS  978
Db  1148  VIRKEQNEVGSTIRTRG-NLSLNAKGDIRIRAAEVGSEQGRLKLAAGRDIKVEAGKAHTE 1206

Qy   979  TPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDLQSKLETIDLG 1038
Db  1207  TEDALK------------YTGRSG--------GGIKQKMTRHLKNQNG-----QAVSGTLD-- 1242

Qy  1039  EHGAELLNDMKRFRDELEQSATRSVTVLG-----QHQGVLKSNGEI-------NSEF 1083
Db  1243  --GKEII------LVSGRDIITVTGSNIIADNHTILSAKNNIVLKAAETRSRSAEM 1289

Qy  1084  KPSPGKALVQSFNVN--RSGQDLSKSLQQAVHATPSAESKLQSMLG--------HFVS 1132
Db  1290  NKKEKSGLMGSGGIGFTAGSKKDTQTNRSETVSHT----ESVVGSLNGNTLISAGKHYTQ 1345

Qy  1133  AGVDMSHQKGEIPLGRQRDPNDKTALTKSRLLDTVTIGELHELADKAKLVSDHKPDADQ 1192
Db  1346  TGSTISSPQGDV------------GISSGKISIDAA------ 1369

Qy  1193  IKQLRQQFDTLREKRYESNPVKHYTDMGFT-------HNKALEANYDAVKAFINAFKKEH 1245
Db  1370  ------QNRYSQESKQVYEQKGVTVAISVPVVNTVMGA-VDAVKA-------- 1407

Qy  1246  HGVNLTTRTVLESQGSAELAKKLKNTL-LSLDSGESM------SFSRSYGGGVSTVFVPTL 1299
Db  1408  --VQTVCKSKNSRVNAMAAAANALNKGVDSGVALYNAARNPKKAAGQGIS----- 1454

Qy  1300  SKKVPVPVIPGAGITLDRAYNLSFSR--------TSGGLNVSFGRDGVSGNIMVATG 1349
Db  1455  --VSVITYGEQKNTSESRIKGTQVQEGKITGGGKVSLIASGAGKDSRITITG 1503

Qy  1350  HDVMPYMTGKKTSAGNASDWLSAK----------HKISPDLRIGAAVSGTLQGTLQNSLKFKL 1402
Db  1504  SDVYG-GKGTRLKAENAVQIEAARQTHQERSENKSAGFNAGVAIA-----INKGISF- 1554

Qy  1403  TEDELPGFIHGLTH--------GTLTPAELLQKGIEHQMKQGSKLITFSVD- 1444
Db  1555  --GFTAGANYGKGYGNGDETAYRNSHIGSKDSQTAIESGGDTVIKGGQLKGKGVGV 1608

Qy  1445  TSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTSASNN 1504
Db  1609  TAESLHIESLQDTAVFKGKQENVSAQVTVGYGFS----VGGSYNRSKSSSDYASVNEQSG- 1664

Qy  1505  RPTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSISLEL 1564
Db  1665  --IFAGGDGYRIRVNGKTGL------VGAAVVSDADKSKNL--L 1698

Qy  1565  KRAE----PVTSNDISELTSTLGK-----HFKDSATTKMLAALKELD----DAKPAEQLHIL 1613
Db  1699  KTSEIWHKDIQNHASAAAASALGLSGGFSYSPKPTSGQYSTKKEAEIGKIGGKPVSLMRFD 1758

Qy  1614  QQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVE 1673
Db  1759  Q--VSAKDDELNEKYRSER-------------IEKGETFKEANLNQ---NNAGGLK 1796

Qy  1674  L-----LHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDG-- 1725
Db  1797  FGLKQNDIHSNDKYAL--AKMGLGNLLGN--------AKESESRQSITRSV--ISEGDW 1843

Qy  1726  --LREQTEKAILDGKVGREEV--GVLFQDRNNLRVKSVSVSQVSKSEGFNTPALLGTS 1781
Db  1844  QIASAQGRKNIAGIEKGTSSAHKALAKADREGL-LKEVELNRDVAKE-FINETLIGGIA 1900

Qy  1782  NSAAMSM--ERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMK 1837
Db  1901  DEAYRSQFIAEHRLMT-FKMDENGEPIEDKQLEEDI---NKQFDNSV-KLKKEFASFK 1953
```

```
Db  1495 LKVNADVVNNAVGEQSAIAGKNGVALQVGGTQLTGGEIRSQOGKVELGGSQVSQQDVN 1554
            :   :   :  :
Qy  1663 LSRINNDGIVELLHKHFDAALPASSAKRLGEMMNDPALKDIIKQLQSTPFSSASVSMEL 1722

Db  1555 GQRYQGGGRVD---------AAATVGGLLGG--AAKQSV--AGNVPFASGHASTQQ 1597

Qy  1723 KDGLREQTEKAILDGK 1738
Db  1598 AD-----AKAGVFSGK 1608


RESULT 15
T30336
nuclear/mitotic apparatus protein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T30336
R;Merdes, A.; Ramyar, K.; Vechio, J.D.; Cleveland, D.W.
Cell 87, 447-458, 1996
A;Title: A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle
A;Reference number: 220828; MUID:97053784
A;Accession: T30336
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-2253 <MER>
A;Cross-references: EMBL:Y07624; NID:g1514670; PIDN:CAA68905.1; PID:g1514671
C;Genetics:
A;Gene: NuMA

Query Match                 2.3%;   Score 220.5;   DB 2;   Length 2253;
Best Local Similarity   18.6%;      Pred. No. 0.013;
Matches 400;   Conservative 300;   Mismatches 718;   Indels 737;   Gaps 101;

Qy    61 STAADGISAAHQQKK------SFSLRGCLGTKKFSRS--APQGOPGTTHSK--------- 103
Db   350 TTEEQELSLSNWQQKONQLESELSGAVGEKKYLEEHNLILQGKISMLEDQLKEMGEIDMP 409

Qy   104 --GATLRDLLARDDGETQHEAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPT 161
Db   410 ETGDCMGDILKLDD--LKQELAVLNTQCLSLKEQI--HQMEE-----EKSTAEVEMEA 458

Qy   162 QQKRHQLNNFGQMR--------QTMLSKMAHPAS----------ANAG 191
Db   459 QKSRFE-SEKGQLQEIVTNLQTSLSEITFQKERLDNEARAQQEHLMCQITTLKLEISKLK 517

Qy   192 DRLQHSPPHIPGSHHEIKEEP------------VGSTSKATTAHADRVEIAQEDD 234
Db   518 SSLVHKDEELKGIHHKVEEFRNEKNQLLENFKMLGGNLIGITQ--------QLESKTK 567

Qy   235 DSEFQQLHQQRLARERENPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQ 294
Db   568 EVDYLREQQOKILCER----------------------------DSTLSTLNEY 593

Qy   295 SMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLK----------- 331
Db   594 KCKKDENSGV----------LNKTVKTLEQDHQTSLSVIEKLKSEKEELASKVQDLDAKMLGLI 647

Qy   332 ---QTLGKDTQHYLAHHASSDGSQOHLLLDNKGHLFDIKSTATSYVLHNSHPGEIKGKLA 388
Db   648 AKCQNLDSENDSQSKSHAATVESLKAQLSEQESQLKIYRKKVSSNELVSEENSKLKDQLL 707

Qy   389 QAGTG----------------SVSVDGKSGKIS----------LGS 408
Db   708 SVEESLRHLREHLEKETKFAASLDADLKRISHLEEEMKKLSESRDEALHNLDEERTAGK 767

Qy   409 GTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSAD 468
Db   768 KIESQLKHLEEEYQKANESLQAKL---AGSC---AAIKQREEERDELSKVVDIWKAKY 819

Qy   469 --------KDTHSQLSRQADG--KLYALKDNRTLQNLSDNKSSEKL 504
Db   820 GESQQKIAQNSCHMQEQTEELKKTHSDVYQQLEGERSKVLMIE---AKASETKSSQ-- 872
```

```
Db   664 ----------------------------SGGE--------------LSVESKT 674
                                                           :  :
Qy   743 SSIKHGTENVFSLPHVRNKPEPGDALQGLNKDKAQAMAVIGVNKYLALTEKGD--- 796
Db   675 GNI----NVKA--------AERQONIDEQKTALTVNGYAK-----EAGDKQYRAG 712

Qy   797 --IRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFHQPREAWQ 854
Db   713 LRIEHTRDSEKTTRTENSASSLS--GGSVKLK-----------AEKDVTFSGSKLVAD 757

Qy   855 NGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLA 914
Db   758 KGDASVSGNKVSFLAADDKTAS----------------------NTEQTKIGGGFYY------- 792

Qy   915 VGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRA---Y 971
Db   793 --TGG---IDKLGSGVE-----------------------AGYENNKTQAQSSKAITSGSDV 826

Qy   972 AFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDLQSK 1031
Db   827 KGNLTINARDKL------TQQGAQ------HSVGGAY--QENAAGVDHLAAADTASTT 870

Qy  1032 LETLDLGEH-GAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGKA 1090
Db   871 TTKTDVGVNIGANV--DYSAVTRPVERAVGKAAKL--DATGVINDIGGIGA---PNVG-- 921

Qy  1091 LVQSFNVNRSGQDLSK--SLQQAVIAHTPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGR 1148
Db   922 ---LDIGAQGGSSEKRSSSSQAVVSS-------------VQAGSIDINAKGEV--------- 958

Qy  1149 QRDPNDKTALTKSRLLLDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRY 1208
Db   959 -RDQGTQYQASK--------GAVNLTAD------SHRSEAAANRQDEQSRDT------R 996

Qy  1209 ESNPVKHYTDMG----------------FTHNKALEANYDAVKAFINAFKKE--HHGVN 1249
Db   997 GSAGVRVYTTTGSDLTVDAKGEGGTQRSNSSASQAVTGSIDAANGINVNVKKDAIYQGTA 1056

Qy  1250 LT---TRTVLESQGSAEL-----AKKLKNTLLSLDSG---ESMSFSRSYGGGV----- 1291
Db  1057 LNGGRGKTAVNAGGDIRLDQASDKQSESRSGFNVKASAKGGFTADSKNFGAGFGGGTHNG 1116

Qy  1292 ----STVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFG----------- 1335
Db  1117 ESSSSTAQVGNISGQGGVELKAGRDLTLQG------TDVKSQGDVSLSAGNKVALQAAESTQ 1172

Qy  1336 --RDGGVSGNIMVATGHDVMPYMTGKKTSAGNASD----------WLS 1371
Db  1173 TRKESKLSGNIDLGAGSSDSKEKTGGNLSAGGAFDIAKVNESATERQGATIASDGKVTLS 1232

Qy  1372 AKHKISPDLRI-GAAVSG---TLQ----GTLQNSLKFKLTEDELPGFIHGLTHGTLTPAE 1423
Db  1233 ANGKGDDALHLQGAKVSGGSAALEAKNGGILLESAKNEQHKDN--------------- 1275

Qy  1424 LLQKGIEHQMKGQSKLTFSVDTSANLD----------LRAGINLNEDGSKPN-GVTA 1469
Db  1276 -WSLGIKANAKGGQ--TFNKDAGGKVDPNTGKDTHTLGAGLKVGVEQQDKTTHANTGITFA 1332

Qy  1470 -RVSAGLSASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGV----------GAGA 1516
Db  1333 GDVTLNSGKDTRLAGARVDADSVQGKVGGDLHV-ESRKDVENGVKVDVDAGLSHSNDPGS 1391

Qy  1517 NLTAALGVAHSSTHEGK----PVG---IFPAFTSTNVSAALALDNRTSQSISELKRAE 1568
Db  1392 SITSKLSKVGTPRYAGKVKEKLEAGVNKVADATTDKYNSVARRLDPQQDTTGAVSFSKAE 1451

Qy  1569 --------PVTSNDISEL------TSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQH 1616
Db  1452 GKVTLPATPAGEKPQGPLWDRGARTVGGAVKDSIT--------GPAGR---QGH 1494

Qy  1617 FSA-----KDVVGDERYEAVRNLKLVIRQQA------ADSHSMELGSASHSTTYNN 1662
```

```
Qy    505  VDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPE-------------------SHISLSLHF  551
                :||              ::        : :||                             ::|| ||
Db    873  LEKINQLEGELSAANACIKEREAEEK-KLVSALHSAEEKLKIAYQGESERLSHLETAL--       929

Qy    552  ADAHQGL--------LHGKSELEAQSVAI---SHGRLVVADSEGKLFSAAIPKQGDGN  598
              :|:|               ||:|| |||                :||   ||  |
Db    930  SNAKQDLDCLAKELSDEKYKKAEFEAMVKLKEQNSERIASLESELK-NSLAVVKERKCE  988

Qy    599  ELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLN-ALVKNNFRQQHACPLGNDHQFHPG  656
             : ||                :           |||  :    :::||
Db    989  SEKLSGEVEHL--------KRQL------DDSSQKHKEALAQKNIEIKQ--LINAKE---  1029

Qy    657  WNLTDALVIDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQ  716
                ||                  ::   |                        |||
Db   1030  -KATSDLAIKSEMG-----AQLQKAVDTHKSEFSALQ                        1062

Qy    717  LKKGLDGAAYLLKDGEVKRLN------INQSTSSIKHGTENVFSLPHVRNKPEPGDA  767
              || ||::|| |||:||||||         ::|| |            :| :     |
Db   1063  LSRSLDLLA--LKEGEVERLNKEAALRQEEIQQQOTITKLTEEETALAALKDKV----A  1116

Qy    768  LQGLNKDDKAQAMAVIGVNKYLA----LTEKG---------DIRS------FQIK  803
              ||:                      : :||          ||:|        ||||
Db   1117  LQ---EKEIKQQVOATKGAEKEMAKLKSVISEKSKRIECLEQDIQNQKRDLSCIQEDHQSK  1174

Qy    804  PGTQQ---------LER----------------PAQTLSREGIS-----------  822
                |:||          ||:           |  ::::|  |
Db   1175  LGESGQLQALIADLEKKCKEQKELICEAQNKAAEKATLASEKASVSERQLEGIQALEIEI  1234

Qy    823  GELKDIHVDHKQNL---YALTHE--------GEVFH--QPREAWQNGAESSSWHKL---  865
             ::|                              |                       |
Db   1235  GKERQKACDLQKQLELSWAVQEEKETELQALKKELFHKVQELEQSQTSFTDSSGEALLYL  1294

Qy    866  ----------ALPQSESKLKSLDMSHEH-----------KPIATFE----DGSQHQL  897
                                ::   ::                  :          ||   |
Db   1295  SEAQERQQALTEAKEQAEQYQKEIEMKNKEVNSLQAEIKILSSKVTTNEEVSVDFEQRLL  1354

Qy    898  KAGGWHAYAAPERGPLAVGTSGSQTVFNRLM---------QGVKGKV----IPGSG  940
            :|                                 : ||
Db   1355  KETSKSKSAKLEEKMQKLHMELEAS--FKELLEKNCAIDCLTTEAQNLKGEADQQRMAVDS  1411

Qy    941  LTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGL  1000
                    :||          :     :  :|:
Db   1412  LQQKLSSKA---------ETNHTLQQEIQAWQ---------KNCAEKEQQICSLQQNL  1451

Qy   1001  KPLYEMQGALIKQLDA--HNVRHNAPQPDLQSKLETLDLGEHGAELLNDMK---RFRDEL  1055
             :|:|:  |  |      | :   : :::       |  :|            ||||:|
Db   1452  K----SNQSLLEEFASLKHSYQEIIAERDLMQE------KHQEELLSHKKLTERFQAEL  1500

Qy   1056  EQSATRSVTVLGQHQGVLKS--NGEIN-SEFKPSPGKALVQSFNVNRSGQDL--SKSL  1108
             ||                   : :|                |||       : : |
Db   1501  EKAKEDMTEIV----LLKEKLHNQELQLHKFQSENSYSLTQISHLQQVNSQLLGANQSL  1555

Qy   1109  QQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGR------QRDPNDKTALTKSR  1162
             |                       :       :|               :|  |
Db   1556  SQISDQGAKKLESEMSTL--------KEQHKEEMKTLRLQYEKTLREGNKQ--VQETS  1603

Qy   1163  LILDTVTIGELHELADKAKLVSDHKPDADQ--------IKQLRQQFDTLR-EKRYE  1209
             :  :|                         :            :|        ||:|
Db   1604  LQLETVTSKYDHV--KSKVLKDQKTFQEEKQRLLLQVQELNKQLSQQEKTIRSQQQKLK  1660

Qy   1210  SNPVKHYTDMGFTHNKALE-------------------ANYDAVKAFINAFKKEHHGVNLTTR  1253
             | :                                  |
Db   1661  QREGETHEEADKSHKRVLELESQLEQQTQAVEHYKAQMEKAKVHYDAKKKQNQELSEELQ  1720

Qy   1254  TVLESQ---------GSAELAKKLKNTLLSLDSGESMSFSRSRYGGGVSTVFVPTLSK  1301
              :  :|                 ::
Db   1721  SHIKQQEHLSKENADLKAESEQLHKELQHSLLQSKEVEQN----------CKNLSN  1766

Qy   1302  KVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGRPGGVSGNIMVAT----GHDVMPY  1355
             |                                                :|      |
Db   1767  RV----------RSLEASWSTLTGQL------RD--PGKFQLATAFHEESGHFCAPR  1805
```

```
Qy   1356  MTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKFLTEDELPGFIHGLT  1415
               :|                  :|| |   :|                :|   :|
Db   1806  QT--RSHADVSID--------SLDLSFG------EDQLLNSTSERCNEEPATSSVHASS  1848

Qy   1416  HGTLTPAELLQK--GIEH---QMKQGSKLTFSVDTSANLDLRAG-----------  1454
            :|                    ||
Db   1849  PDSLISGQLPKKVESLESLYFTPITPRAQSKLESSIGDLSLDSSKKTQSARRRTMQI  1908

Qy   1455  INL-----NEDGSKPNGV-------------TARVSAGLSASANLAAGSRER  1488
            |:|                           |
Db   1909  INITMTKKTKEEPESANTSFYSLRSAPSYQSLHLQNPRRAGRPPAAISAPALASLFSQES  1968

Qy   1489  STTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVA--HSSTHEGKPVGIFPAFTSTNV  1546
                 |              :|  |    || | |      :: |  | || :
Db   1969  LAKTEHFSSDDSL-NNLPGYQHPTRRSARLSQTGGRSSFYMSTCQDEP---DPQDDWTRI  2024

Qy   1547  SAALALDNRTSQ--SISLELKRAEPVTSNDISELTSTLGK----------HFKDSA  1590
             :                                                  : |
Db   2025  -AELQARNKTCPPHLKTSYPLESRPSIFSSTITDEEVKLGDPKETLRRATLLPGQIQDSM  2083

Qy   1591  TTKMLAALKELDDAKP-AEQL--HILQQHFSAKDVVGDERY----EAVRNLKKLVIRQO  1642
                                                                |:
Db   2084  TSTRRQTL-----AVPGAEHLKGHNISTRQQMKRVSEESHYGPDTPEAKTATCFPRPMT  2138

Qy   1643  AADSH-------SME-LGSASHS-----------TTYNNLSRINNDGIVELLHKHFDAAL  1683
            ||:|                                    |    ||  ||||||:|
Db   2139  PKDKHDTRRVSTMESKGSSSHQOAQPTRRQATAFSIFNTPKKLGNSLLKRGLNK---KTT  2195

Qy   1684  PASSAKRLG---EMMNNDPALKDIIKQ-LQSTPFSSASVSMELKDGLREQTEK  1732
                                  ::
Db   2196  PKNSPRGRGANGSTSSTSNKPSHLSLRKSPSQRSPRVSTAKSPRASNKLFERKQQ  2250
```

Search completed: June 5, 2001, 18:18:11
Job time: 237 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 5, 2001, 18:14:44 ; Search time 30.94 Seconds
                                        (without alignments)
                                      898.453 Million cell updates/sec

Title: US-09-596-784-2
Perfect score: 9448
Sequence: 1 MELKSLGTEHKAAVHTAAHN.........NPQVASALTDLKKEGLEMKS 1838

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 296979 seqs, 15124130 residues

Total number of hits satisfying chosen parameters: 296979

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Pending_Patents_AA_New:*
               1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
               2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
               3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
               4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
               5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
               6: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	204.5	2.2	1992	4	US-08-945-567C-3	Sequence 3, Appli
2	198.5	2.1	3170	6	US-60-248-505-909	Sequence 909, App
3	196.5	2.1	2031	5	US-09-739-449-10549	Sequence 10549, A
4	195	2.1	1879	1	PCT-US00-35017A-1265	Sequence 1265, Ap
5	194	2.1	2263	1	PCT-US01-04098A-1662	Sequence 1662, Ap
6	193.5	2.0	2515	1	PCT-US01-04098A-3914	Sequence 3914, Ap
7	193.5	2.0	2515	1	PCT-US01-04098A-3915	Sequence 3915, Ap
8	190	2.0	5447	1	PCT-US01-03782A-284	Sequence 284, App
9	188.5	2.0	1423	5	US-09-196-296B-2	Sequence 2, Appli
10	188.5	2.0	5373	1	PCT-US01-03782A-96	Sequence 96, Appli
11	184.5	2.0	1833	4	US-08-945-567C-4	Sequence 4, Appli
12	167	1.8	2189	1	PCT-US01-04098A-1679	Sequence 1679, Ap
13	164	1.8	1228	5	US-09-739-449-8806	Sequence 8806, Ap
14	163.5	1.7	1654	1	PCT-US01-04098A-3647	Sequence 3647, Ap
15	163	1.7	1242	5	US-09-739-449-13178	Sequence 13178, A
16	162	1.7	1668	1	PCT-US01-04098A-1946	Sequence 1946, Ap
17	162	1.7	1715	1	PCT-US01-04098A-1947	Sequence 1947, Ap
18	160.5	1.7	1563	1	PCT-US01-04098A-1381	Sequence 1381, Ap
19	158	1.7	2308	5	US-09-644-293-2	Sequence 2, Appli
20	153.5	1.6	1203	1	PCT-US01-04098A-1926	Sequence 1926, Ap
21	152.5	1.6	1247	1	PCT-US01-04098A-1393	Sequence 1393, Ap
22	152	1.6	2128	6	US-60-248-505-927	Sequence 927, App
23	148.5	1.6	2189	1	PCT-US01-04098A-3589	Sequence 3589, App
24	146.5	1.6	2161	1	PCT-US01-04098A-1621	Sequence 1621, Ap
25	145	1.5	1441	6	US-60-285-697-10	Sequence 10, Appl
26	145	1.5	1960	1	PCT-US01-04098A-1516	Sequence 1516, Ap
27	144.5	1.5	2399	5	US-09-739-449-11472	Sequence 11472, A
28	142.5	1.5	1872	1	PCT-US01-04098A-1822	Sequence 1822, Ap
29	142.5	1.5	3418	5	US-09-421-124-44	Sequence 44, Appl
30	142.5	1.5	3418	5	US-09-421-124-44	Sequence 44, Appl
31	142	1.5	1041	1	PCT-US01-04098A-3234	Sequence 3234, Ap
32	142	1.5	2940	1	PCT-US01-04098A-1823	Sequence 1823, Ap
33	140	1.5	1348	1	PCT-US01-04098A-3173	Sequence 3173, Ap
34	140	1.5	1359	1	PCT-US01-04098A-1205	Sequence 1205, Ap
35	139.5	1.5	1368	5	US-09-739-449-11090	Sequence 11090, A
36	139	1.5	1976	5	US-09-739-449-9987	Sequence 9987, Ap
37	139	1.5	2063	1	PCT-US01-04098A-1372	Sequence 1372, Ap
38	139	1.5	2476	5	US-09-824-574-7	Sequence 7, Appli
39	138.5	1.5	1963	1	PCT-US01-04098A-3484	Sequence 3484, Ap
40	138	1.5	1044	6	US-60-248-505-860	Sequence 860, App
41	138	1.5	2136	1	PCT-US01-04098A-1357	Sequence 1357, Ap
42	138	1.5	2816	1	PCT-US01-10484-145	Sequence 145, App
43	137.5	1.5	1047	6	US-60-248-505-1184	Sequence 1184, Ap
44	137	1.5	648	5	US-09-813-408-26	Sequence 26, Appl
45	137	1.5	699	1	PCT-US01-04098A-1818	Sequence 1818, Ap

ALIGNMENTS

RESULT 1
US-08-945-567C-3
; Sequence 3, Application US/08945567C
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF
; TITLE OF INVENTION: MORAXELLA
; FILE REFERENCE: 1038-745 MIS
; CURRENT APPLICATION NUMBER: US/08/945,567C
; CURRENT FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 08/431,718
; PRIOR APPLICATION FILING DATE: 1995-05-01
; PRIOR APPLICATION NUMBER: 08/478,370
; PRIOR APPLICATION FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/621,944
; PRIOR APPLICATION FILING DATE: 1996-03-26
; PRIOR APPLICATION NUMBER: PCT/CA96/00264
; PRIOR APPLICATION FILING DATE: 1996-04-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1992
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-08-945-567C-3

  Query Match   2.2%; Score 204.5; DB 4; Length 1992;
  Best Local Similarity   19.0%; Pred. No. 0.0011;
  Matches 318; Conservative 221; Mismatches 606; Indels 531; Gaps 79;

Qy   469 KDT-HSQLSRQ--------ADG-KLYALKDNRTLQN------LSDNKSSEKLVDKIKS 510
                                                      ||| :|
Db   15 KDTKHIAIGEQNQPRRSGTAKADGDRAIAIGENANAQGGQAIAIGSSNKT-------VNG 67

Qy   511 YSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISL--SLHFADAHQGLLHGKSELEA 567
               |:|| :| | :| || :: :|:   ::
Db   68 SSLDKIG-----TDATGQESIAIGGDVKASGDASIAIGSDDLHLLDQHGNPKHPKGTLIN 122

Qy   568 QSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDD 627
                          |:|: |:|:
Db   123 DLI---NGHAVLKEI------------RSSKDNDVKYRRTTAS----GHASTAVGAMSYA 163

Qy   628 HGQLNALVKNNFRQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMG--- 685
                                                            || :|
Db   164 QGHFS----NAF--------------GTRATAKSAYSLAVGLAAT-AEGQSTIAIGSD 202

us-09-596-784-2.rapn

Wed Jun 6 10:19:11 2001

```
Qy   686 ----HIGSLALQEG--KLHYFDQLTKGWTGAESDCKQL------KKGLDGAAYLLKDGE  732
             :!:!!!   ::!!!!  :!    :!
Db   203 ATSSSLGAIALGAGTRAQLQGSIALQOGSVVTQSDNNSRPAYTPNTQALDPKFQATNNTK  262

Qy   733 VKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDK--AQAMAVIGVNKYL   789
             !:!  ::!!!!  !!   !!       !
Db   263 AGPLSI--GSNSIKRKIINVGA---------GVNKTDAVNVAQLEAVVKWAKER       305

Qy   790 ALTEKGDIRSFQIKPGTQQ------LERPAQTLSREGISGELKD-IHVDHKQNLYALT   840
             !!   !!!   ::!
Db   306 RITFQGDDNSTDVKIGLDNTLTIKGGAETNALTDNNIGVVKEADNSGLKVKLAKTLNNLT  365

Db   841 HEGEVFHQPREAWQNGAESSSWHKLAL---------PQSESK----LKSLDMSHE     882
             ::!     !!          :!!
Qy   366 EVNTTLNATTTVKVGSSSSTAELLSDSLTFTQPNTGSQSTSKTVYGVNGVKFTNNAET   425

Db   883 HKPIAT--------FEDG------SQHQLKAG-----GWHAYAAPERGPLA        914
             :!!            !!!
Qy   426 TAAIGTTRITRDKIGFARDGDVDEKQAPYLDKKQLKVGSVAITIDNGIDA-GNKKISNLA  484

Qy   915 VGTSGSQTVFNRLMQGVKGKVIPGSGLTV----KLSAQTGGMT-------GAE       956
             !:     !:!:!!!  :!!:!
Db   485 KGSSANDAVTIEQLKAAKPTLNAGAGISVTPTEISVDAKSGNVTAPTYNIGVKTELNSD   544

Qy   957 GRKVSSKFSERIRAYAFNPTMSTPRPI-----------------              983
             !:!!:!   :!:!!
Db   545 G--TSDKFS--VKGSGTNNSLVTAEHLASYLNEVNRTADSALQSFTVKEEDDDANAITV   600

Qy   984 -----KNAAYATQHGWQGREGLKPLYEMQGALIKQL--------DAHNVRHNA      1023
              !!!:   :!!     !!:!!
Db   601 AKDTTKNAGAVSITLKLKGNGLITVATKKDGTVFTGLSQDSGLTIGKSTLNNDGLTVKDTN 660

Qy  1024 PQPDLQS--KLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQVLKSSNGEIN  1080
             :!:!:   :!!!::    :!!
Db   661 EQIQVGANGIKFTNVNGSNPGTGIANTARITRDKIGFAGS-----DGAVDTNKPYL     711

Qy  1081 SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ 1140
             :!!:!     !!!!
Db   712 DQDKLQVGNVKITNTGINAGGKATT----GLSPTLPSI--------ADQSSR      751

Qy  1141 KGEIPLGRQRDPNDKTALTKSRLILDT---------VTIGEL               1173
             !:!    :!!!   !:!
Db   752 --NIELGNTIQDKDKSNAASINDILNTGFNLKNNNNPIDFVSTYDIVDFANGNATTATVT  809

Qy  1174 HELADK-AKLVSD---------HKPDADQIKOLRQQFDTLREKRYESNPVKHYTDMGFTHN 1224
             !:!:!:!  :!!       !:!    :!
Db   810 HDTANKTSKVVYDVNVDDTTIHLTGTDDNKKLGVKTTKLNKTSANGN-----TATNFNVN  864

Qy  1225 KALE--ANYDAVKAFINAFKKEHHGVNLITRTVLESQGSAELAKLKNTLLSLDSGESM   1281
             !!    ::!!   !!!:!
Db   865 SSDEDALVNAKDIAEENLNTLAKEIHTTKGTADTALQT---FTVKKVDENNNADDANAIT  920

Qy  1282 SFSRSYGGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFS-RTSGGLNV-SFGRDG  1338
             :!!!:!   !!!!!  :!:!     !!
Db   921 VGQKNANNQVNTL--TLKGE----NGLNIKTDKNGTVTFGINTTSGLKAGKSTLNDG    971

Qy  1339 GVS-------GNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHHISPDLRIGAAVSGTLQGT 1393
             !!!        !!!:  :!!
Db   972 GLSIKNPTGSEIQVGADGVKFA---------KVNNNGVVGAGIDGTT---            1010

Qy  1394 LQNSLKFKLTEDELPGFIHGLTHGLTLPAE--LLQKGIEHOMKOGSKLTFSVDTSANLDL 1451
             !:!!    :!!!   !!!:!        ::!
Db  1011 ------RITRDEI-GPTG--TNGSLDKSKPHLSKDGIN--AGGKIT-------NI     1048

Qy  1452 RAG--INLNE----DGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTSASNNRP 1506
             !!!  ::!!     !!  !!!!
Db  1049 QSGEIAQNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLHEFSVADEQGNNFTVSNPYS  1108

Qy  1507 TF--------LNCVGAGANLTAALGVAHSSTHECKPVGIFPAFTSTNVSAALALD       1553
             !:!          :!!:!!!      ::!
Db  1109 SYDTSKTSDVITFAGENGITTKVN----KGVVRVGIDQTKGLTTPKLLTVGNNNGKGIVID 1164
```

```
Qy  1554 NRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSA----TTKMLAALKELDDAKPAEQL 1610
             !!   :!:!!           !:!
Db  1165 SQNGQ----------------------NTITGLSNTLANVTNDKGSVRTTEQGNIIKDEDKTRAASIV 1210

Qy  1611 HILQQHFS------AKDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTYNNLS    1664
             !:!:!!       !:!      :!
Db  1211 DVLSAGFNLQGNGEAVDFV--STYDTV-------NFADGN-                     1253

Qy  1665 RIN------NDGIVELLHKHF                                          1686
             !!!       !:!:!
Db  1254 KTSKVVYDVNVDDTTIEVKDKKLGVKTTLTSTGTGANKFALSNQATGDALVKASDIVAH  1313

Db  1687 ------SAKRLGEMNN----DPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKA      1733
             :!!!    :!!!
Db  1314 LNTLSGDIQTAKGASQANNSAGYVDADGNKVIYDSTDNKYYQAK---NDGTVDKTKEV   1368

Qy  1734 ILLDGKVGREEV--GVLFQDRNNLRVKSVSVSQSV--SKSEGFNTP-ALLLGTSNSAAMS 1787
             !:!:!     :!:!!      :!!
Db  1369 AKDKLVAQAQTPDGTLAQ----MNVKSVINKEQVNDANKKQGINEDNAFVKGLEKAASDN 1424

Qy  1788 MERN---IGTIN----FKYGGDQNTPRR------FTLEGGIAQANPQVASALTD       1828
             !:!    !!!:!   :!!!
Db  1425 KTKNAAVTVGDLNAVAQTPLTFAGDTGTTAKKLGETLIIKGGQTDTN------KLTD   1475
```

RESULT 2
US-60-248-505-909
; Sequence 909, Application US/60248505
; GENERAL INFORMATION:
;   APPLICANT: Beasley, Ellen
;   TITLE OF INVENTION: ISOLATED HUMAN G-PROTEIN COUPLED
;   TITLE OF INVENTION: RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR
;   TITLE OF INVENTION: PROTEINS, AND USES THEREOF
;   FILE REFERENCE: c1000918
;   CURRENT APPLICATION NUMBER: US/60/248,505
;   CURRENT FILING DATE: 2000-11-15
;   NUMBER OF SEQ ID NOS: 1998
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 909
;   LENGTH: 3170
;   TYPE: PRT
;   ORGANISM: Human
US-60-248-505-909

```
Query Match             2.18%;  Score 198.5;  DB 6;  Length 3170;
Best Local Similarity   18.5%;  Pred. No. 0.0063;
Matches  422;  Conservative 289;  Mismatches 778;  Indels 791;  Gaps 101;

Qy    4 KSLGTEHKAAVHTAAHNPVGH-GVALQQGSSSSPQNAAASLAAEGKNRGKMPRIHQPS    61
            :!:!   !!!!:!!      :!
Db  807 KDVGSEGK---HVRPQIP-GRKWGGFLGRRSKWDGPQNKKDK---EGVLLSKAEKTGEPQ  859

Qy   62 TAADGISAAHQQKKSFSLRGCLGT----KKFSRSAPQQQPGTTHSK-------       103
            !   :!:!!:!      !!
Db  860 T--------QMEKTSQVQELGDDLRMGEKAGELRSTTGKAGDSGLDGSRAHCEKPPCK    910

Qy  104 --------GATLRDLLARDDG---------ETQHEAAAPDAARLTR                132
             :!!!    :!!
Db  911 APEDRWYEAEKVWLAQKDGFTLATVLKPDEGTADLPAGRVRLCIDADKTITEVDEEHVHR  970

Qy  133 SGGVKRRNMDDMAGRPMVKGGS---------GEDKVPTQOKRHQLNNFGQM           174
            :!!!    :!!!
Db  971 ANPPELDQVEDLASLISVNESSVLNTLLQRYKAQLLHTCTGPDLIVLQPRCPSVPSAGKP 1030

Qy  175 RQTMLSKMAHPASANAGDRLQHSPPHIPG--------SHHEIKEEPVGSTSKAT       220
            :!!:!   !!!:!
Db 1031 KASELSTLPRDSSGSL------EIPGRDIFIDPDSNOVLIGHPGLRHQNLGKSAKRC   1081

Qy  221 TAHADRVEIAQEDDSEFQOLHQQRLARERENPQQPPK--LGVATPISARFQPKLTAVA    277
            !!!:!   !:!!!
Db 1082 RIHWLDQPLIPQ------MATWEHRPAVPKGRRDGLPAHIGSMAQRAYAWALL        1127
```

```
Qy   278  ESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAVTLDKGKLQLAPDNPPALNTLLKQTLGKD  337
               :::  ::    ::   :::: ::
Db  1128  --------NQRRDQSIVGRSGAGKT------------------------TCCEQVL--  1151

Qy   338  TQHYLAHHASSDGSQHLLLDKNGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVSV  397
            :   :  ::     :::: ::  :  :::
Db  1152  -EHLVGMAGSVDG--RVSVEKIRATFTVLRAFGSVSMAHS------RSATRFSMVMSL  1200

Qy   398  D-GKSGKISLSGSGTQSHNKTMLSQPGEAHRSLLTGTWQHPAGAARPQGESIRLHDDKIHI  456
          : ::  :  ::      :::::      ::   :::::::
Db  1201  DFNATGRI--TAAQLQTMLLEKSRV----------ARQPEGESNFLVFSQMLA  1241

Qy   457  ----LHPELGVWQSAD-------KDTHSQLSRQADGKLYALKDNRTLQNLSDNKSS  501
               :::   ::        :::     :::
Db  1242  GLDLDLRTELNLHQMADSSSFGMGVWSKFPEDKQKAAAAFAQLQG----AMEMLGISESE  1296

Qy   502  EKLVDKI-----------KSYSVDQRG--QVAILTDTPGRHKMSIM  534
                :             :::           ::
Db  1297  QRAVWRVLAAIYHLGAAGACKVGRKQFMRFEWANTAAEALGCEYEELNTATFKHHLRQII  1356

Qy   535  PSLDASP---------ESHISLSLHFADAHQGLLHGKSELEAQSVA----------ISHGR  576
             ::        :: :::  :::                  ::
Db  1357  QQMTFGPSRWGLEDEETSSGLKMTGVDCVEGMASGQ-ELFAAVVSLINRSFSSHHLSMAS  1415

Qy   577  LVVADSPEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFH----------  625
          :  ::        :      :::          :::
Db  1416  IMVVDSPG---------FQNPRHQGKDRAATFEELCHNYAHERLQLLFYQRTFVSTLQRYQ  1467

Qy   626  ---------DDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNL-------TDALV  664
                  :  :     ::       :::  ::          ::
Db  1468  EEGVPVQFDLPDPSPGTTVAVVDQN-PSQVRLPAGGAQDRLFWVLDEEVHVEGSSDSVV  1526

Qy   665  ID-----------NOLGLHHTNPEPHEILDMGHLGSLAL----QEGKL  697
          ::           :  :::     ::      ::          :::
Db  1527  LERLCAAFEKKGAGTEGTFAECSHKGQOGLGSTIMEPRH--SGQSGSSALRTCEQPLQC  1583

Qy   698  HYFDQLTKGW---------TGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIK----  746
            ::           :::  :  :::       :::
Db  1584  EIFHQL--GWDPVRYDLTGWLHRAKPNLSALDAPQVLQQSKREELRSLFQARAKLPPVCR  1641

Qy   747  -----HGTEN---------VFSLPHVRNKPEPG-----DALQGLNKDDKAQAMAV  782
               :::          :::      :::          :::
Db  1642  AVAGLEGTSOQALQRSRMVRRTFASSLAAVRRKAPCSQIKLQMDALTSMIKRSRLHFIHC  1701

Qy   783  IGVNKYLALTEKGDIRSFQIKP-----------GTQQLERPAQTLSREGISGELKDIH  829
              :  :::         ::            :::    :::
Db  1702  LVPNPVV-----ESRSGQESPPPQPGRDKPGAGGPLALDIPALRVQLAGFHIALR-LH  1754

Qy   830  --VDHKQNLYALTHEGEVFHQP-REAWONGAESSSWHKLALPQSESKLKSLDMSHEHK  884
              ::        ::         ::    :::          ::
Db  1755  RTGYADH-MGLTRFRRQFQVLDAPLLKKLMSTSEGIDERKAV----EELLETLDL--EKK  1807

Qy   885  PIATFEDEGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVK  944
              :::      :::         :::          ::
Db  1808  AVAV--GHSQVFLKAG----------VISRL-EKQREKLVSQS-IVLF  1841

Qy   945  LSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATOHGWQ---------  995
              :::          :::      ::           :::
Db  1842  QAACKGFLSRQEFKKLKIR--RLAAQCIQKNVAVFLAVKDWPW----WQLLGSLQPLL  1893

Qy   996  ---GREGLKPLYEMQGALIKQLD-AHNVRHNAPQPD--LQSKLETLDLGEHGAELLNDM  1048
              :::         ::            ::           :::
Db  1894  SAFIGTEQLRAKEELTLRRKLEKSEKLRNELRQNTDLLESKI-------ADITSDL  1944

Qy  1049  --KRFRDE-----LEQS--ATRSVTVL-GQHQGVLKSNGEINSEFKPSPGKALVQSFN  1096
             ::       ::      :::           ::
Db  1945  ADERFKGDVACQVLESAERLQAFREVQELKSKHEQVQKKLGDVNKQLEEAQQKIQLNDLE  2004

Qy  1097  VNRSG---------QDLSKSLQQAVHA--TPPSAESKLQSMLGHFVSA---GV  1135
             :::          ::         :::           :::
Db  2005  RNPTGGADEWQMRFDCAQMENEFLRKRLQQCEERLDSELTARKLEQKLGELQSAYDGAK  2064

Qy  1136  DMSHQKGE--------------IPLGRQRDPNDKTALFKSRLIL--  D  1166
              :::            :::
Db  2065  KMAHQLKRKCHHLTCDLEDTCVLLENQQSRNHELEKKQKKFDLQLAQALGESVFEKGLRE  2124

Db  1167  TVT-------IGELH-------------ELADKAKLVSDHKPD--  1189
Db  2125  KVTQENTSVRWELGQLQQQLKQKEQEASQLKQQVEMLQDHKRELLGSPSLGENCVAGLKE  2184

Qy  1190  --------ADQIKQLRQQF----------DTLREKRY  1208
              :::         ::           :::
Db  2185  RLWKLESSALEQQKIQSQQENTIKQLEQLRQRFELEIERMKQMHQKDREDQEEELEDVRQ  2244

Qy  1209  ESNPVKHYTDMGF------------THNKA-LEA---NYDAVKAFINAFKKEH  1245
              :::          ::            :::
Db  2245  SCQKRLHQLEMQLEQEYEEKQMVLHEKQDLEGLIGTLCDQIGHRDFDVEKRLRRDLRRTH  2304

Qy  1246  --HGVNLTRTVLESQGSA---ELAK--KLKNTLLSLDSGESMSFSRSYGGVSTVFVP  1297
              :::          ::            :::
Db  2305  ALLSDVQLLLGTMEDGKTSVSKEELEKVHSQSTLLALKKGTSSGMWGLVNKAVVTLPVQ  2364

Qy  1298  TLSK------KVPVPVIP-GAGITLDRAYNLSFSRTSGGLNVSFGRDGVSGNIMVAT  1348
              :::          ::            :::
Db  2365  VISSYRGGSCPQVPPLVLPVDTVGIWLEQ--------------------  2393

Qy  1349  GHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELP  1408
              :::          ::            :::
Db  2394  ----------SEAKCEEALKTQKVLTADLE--------SMHSELENMTRNKSLVDE--  2431

Qy  1409  GFIHGLTHGLTLTPAELLQKGIEHQMKQGSKLTFSVDTSANLDLRAGINLNEDGSKPNGVT  1468
              :::          ::            :::
Db  2432  ---QLYRLQFEKADLLKRIDEDQ----------DDLNELMQKHKDLI  2465

Qy  1469  ARVSAGL----SASANLAAGSRERSTTSGQFGSTTS----ASNNRPTFLNGVGAGANLT  1519
              ::           ::
Db  2466  AQSAADIGQIQELQLQLEAKKEKHKLQEQAGSCTCLCPGIRDNRCPIDFSHTYSCRWLR  2525

Qy  1520  AALGVAHSSTHEGKPVGIFPAFTSTNVSAALA-LDNRTSQSISLELKRAEPVTSNDISEL  1578
              :::          ::            :::
Db  2526  CA----SSTWNSPP-WVEP--SSAGRKAVICDLENKT-EFQKVQIKRFEVPISLQPAGF  2576

Qy  1579  TSTLGKHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEAVR-NLKKL  1637
              :::          ::            :::
Db  2577  TTVLVIRLRDS---LIKMGEELSQAATSES----QQRESSQ--YYQRRLEELKADMEEL  2626

Qy  1638  VIROQAADSHSMELG---------SASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKRL  1691
              :::          ::            :::
Db  2627  VQREAEASRRCMELEKVEELAAVRQTLQTDLE----------TSIRRI  2665

Qy  1692  GEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVGREEVGVLFQDR  1751
              :::          ::            :::
Db  2666  AD------LQAALEEVASSDSDTESVQTAVDCG--SSGRKEIIRGHKSRVPVLKMIQIR  2716

Qy  1752  NNLRVKSVSVSQSVSKSEGFNTPALLGT-------SNSAAMSMERNIGTINFKYGQDQNT  1805
              :::          ::            :::
Db  2717  DNVSILS-----SQPEGSLQSWLSCTLSLATDTMRTPSRQSATSSRILSPRINEEAGDTERT  2773
```

RESULT 3
US-09-739-449-10549

us-09-596-784-2.rapn

Wed Jun 6 10:19:11 2001

```
US-09-739-449-10549

Query Match     2.1%;  Score 196.5;  DB 5;  Length 2031;
Best Local Similarity   18.99%;  Pred. No. 0.0038;
Matches  337;  Conservative 249;  Mismatches 714;  Indels 485;  Gaps 73;

Qy   325 ALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIK 384
Db   121 AMNDGIERTIARATELETLVHSEVNALERSYADNELRVRSLVQELTAERDAIYNHAERIR 180
Qy   385 GKLAQACTGSVSVDGKSGKISLGSGTQSHNKTMLSQPGE--AHRSLLTGIWQHPAGAARP 442
Db   181 SSIV-------------GAQEQIKEELSIVGEELSMRIATTG--EAFASMI 216
Qy   443 QGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSS- 501
Db   217 DTRSAALLE----------KSRASTEAMGSLIAAKTENLLQALNSSGSTI 256
Qy   502 ----EKLVDKIKSYSVDORGQOVAILTDTPGRHKMSIMPS-LDASPES------------H 544
Db   257 SNEFDMRLHNLTS-TLDERGEVLL---------ERFAIHASTLDSGVESLNSALEERTRQLN 308
Qy   545 ISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDG------- 597
Db   309 ETLSARSLELNRNIERGQVIGGSLDTVLDKLSTTLEEKGLSFRQSLQSTADDAIMDLDL 368
Qy   598 --NELKMKA--MPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGN- 649
Db   369 RSGLYEERMQATVGQVNSAFDEHVA--QFASAFDQRAGSLDSKLMESLARINETVAGGS 425
Qy   650 ---DHQFHPGWNLTDALVIDNQLGLHHTNPEPHEILDMG----HLGSLALQEGKLHYFD 701
Db   426 EALDTILTGSLERIGSTMTDQSLALATALGTGQEMLENALESRTQAFSDAIGQRTAEITD 485
Qy   702 QLTKGWTGAESDCKQLKKGLDGAAYLLKD----GEVKRLNINQSTSSIKHGTENVFSLPHVR 759
Db   486 AFTNSHAKIDTVLAERSNALFGALSASQDRFDEALASRSLAITGSVSGTAEHLAAM---- 541
Qy   760 NKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLE-RPAQTLSR 818
Db   542 -----LDERAAAINSVVADVERRLTETLETRAAAITGAVSGIEDRISDTL-- 586
Qy   819 EGISGELKDIHVDHKQNLIYALTHEGEVFHQPREAWQNGAESSSWHKLALPQSESKLKSLD 878
Db   587 ESRTAALHDV-VSGAESRIADTLDG-----RTAALSSAISGVEERIA-DTMDSRTLSLD 638
Qy   879 MSHEHKPIATFEDGSQHQL-KAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVI 936
Db   639 MT-----FANVEERLSETLDNRTSALTGIVASAEEKIAGALDSRTATFGDVVAGAETRIA 693
Qy   937 PG-SGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQ 995
Db   694 ETLDGRTAALNAV---VSGAEERIADALDS---RTMALDMTFSGAEEKIAEALDTRTAAL 747
Qy   996 GREGLKPLYEMQGALIKQLDA-HNVRHNAPQ---PDLQSKLETLDLGEHGAELLNDMKRF 1051
Db   748 GELVASAETRIAGALDSRTDSLKTVVSGAEERITDVLDSRTMALDMSFSGVE-----EKI 802
Qy  1052 RDELE-QSATRSVTVLGQH------------------QGVLKSNGEINSEFKPSPGKAL--- 1091
Db   803 TDILDGRTAALKSAVAGVEDRIAGALDSRTAALSGIVSGAEERIAEALDSRTLALDMTIS 862
Qy  1092 -----VQSFNVNRSGQDLSKS-LQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEI 1144
Db   863 GVEERIAEAMDARASSLSLAAAGVGQRLEATAAFTLENALAS-----------GHERLET 910
Qy  1145 PLGRQRDPNDKTALTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLR 1204
Db   911 MLGSQAERIAGSLERNSGLLIEQSVS---------- 935
Qy  1205 EKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAEL 1264

Db   936 --GAANRIENVVEDG--SSRFAQTVEEGVSRLENNLSQSHEEI---RTALD-QRQADL 985
Qy  1265 AKKLKNTLLSLD--SGESMSFSRSYGGVSTVFV----PTLSKKVPVPVIPGAGITL 1315
Db   986 AATLSSATTQMGDMLSEQAMMIGTTVASSASMELSLSETQQDTLQK----AIDGSAATL 1040
Qy  1316 DRAYNLSFSRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNAS--DWLSA 1372
Qy  1041 E---ARLRNSAGDIAVKIGEAAREIGGATDALS---TRIETSIGNVTTRLDETGA 1089
Qy  1373 KHKISPDLRIGAAVSGTLQGTLQNSLK--------FKLTEDELPGFIHG-LT 1415
Db  1090 RIETSLD-ALQTRVGGDL-ANVNNSIEDAGRRFADALEDKTAVFARTSDEAAERITGILD 1147
Qy  1416 HGTLTPAELLQKGIEHQMKQGSKLTFSVDT-SANLDLRAG----------INLNEDGS 1462
Db  1148 EQTTRVADTFE-------NRTSRLAETFDAGTARIDERLGTMDRALTIGLENVNRTIEGK 1200
Qy  1463 KPNGVTARVSAGLSASANL---AAGSRERSTTSG-QFGSTTSASN--- 1503
Db  1201 ASDLAVSLRGAVVSATQNIGDEAARSSALLAKSGSEFAEQVQAQNEAFTKAIEERSGEIV 1260
Qy  1504 ----NRPTFLNGVGAGANLTAALG---VAHSSTHEGKPVGIFPAFTSTNVSAAL- 1550
Db  1261 TRISDAQTRLSGQAAAVAQTFSEAGNIIVNKVAEEAEAVVRSQVGVI-SETLTSVESALDA 1319
Qy  1551 ------ALDNRTSQ--------SISLELKRAEPVTSNDISELTSTLGKHFKD-------- 1588
Db  1320 RGESIRSALDNRTRELNSMLASRSAELSRLIEEKAKPVVEEYATIGREAAEKIVSAAQQS 1379
Qy  1589 SATTKMLAALKELDDAKPAEQLHILQQHFSAKD 1621
Db  1380 AELLSQSNSGMVGMVEQAISDYAAAGTDAASKLVAATRQSTDM---LSETHTAMAD 1432
Qy  1622 VV--GDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRI---NNDGIVE 1673
Db  1433 AVEQGIDKY--ARSGSDAANKLVAATRQSTDMLSQTH---NSMAEIVEQSATNFNAAVE 1486
Qy  1674 LLHKHF--DAALPAS----SAKRLGEMMNNDPALKD------------1703
Db  1487 RAAQGFGAADEALNASATRFSESASQAADMVSSSRLLEGKIDRLSNISGQTLAQVAGIV 1546
Qy  1704 --IIKQ--LQSTPFSSASVSMELKDGLR---------EQTEKAILDGKVGR 1741
Db  1547 GRFEEHSKVLSQASELLNAAQSLVGTLEERQDALRSLSVGLVKRSEEIETAM---R 1600
Qy  1742 EEVGVLFQD---------RNNLRVKSVSVSQSVSKSEGFNTPA---LLLGT 1780
Db  1601 NVVGVVENTLNEAEERSQNVAGNLRDNLQASFSDIGRSLDETEQRARSAAQTMRGALLSA 1660
Qy  1781 SNSAAMSMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASA 1825
Db  1661 GQDASRSIESTLSDAQ-KYSDELVNRLRGGVESSLSEVDNLLGSA 1704
```

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1879)
;   OTHER INFORMATION: Xaa = any amino acid or other as shown in Table 3
PCT-US00-35017A-1265

Query Match       2.1%;   Score 195;  DB 1;  Length 1879;
Best Local Similarity  18.6%;  Pred. No. 0.0041;
Matches 242;  Conservative 206;  Mismatches 500;  Indels 352;  Gaps 57;

Qy   47 EGKNRGKMPRIHQPSTAADGISAAHQQKKSFSLRGCLGTKKFSRSAPGQGPGTTHSKGAT 106
            ::|||    ||   :  :  |    :|:               ||:
Db  832 EEEDRGQQLQAERKKMAQQMLDLEEQLEEEAARQKIQLEKV----------TAEAKIKK 881

Qy  107 LRD-LLARDDGETQHEAAAPDAARLTRSGGVKRRNMDDMAGRPMVKGGSGEDKVPTQQKR 165
           :|    |  ||                        |||
Db  882 LEDEILVMDDQNN-----------KLSKERKLLEERISDLT-----------TNLAEEEEKA 921

Qy  166 HQLNNFGQMRQTMLSKMAHPASANAGDR-LQHSPPHIPGS----HHEI----------- 208
        ||    |
Db  922 KNLTLKNKHESMISELEVRLKKEEKSRQELEKLKRKLEGDASDFHEQIADLQAQIAELK 981

Qy  209 -----KEEPVGSTSKATTAHADRVEIAQEDD-----------SEFQQ-LHQQRLARERE 251
Db  982 MQLAKKEEEL----QAALARLDD-EIAQKNNALKKIRELEGHISDLQEDLDSERAAR---- 1033

Qy  252 NPPQPPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPLKPQSMLKGSGAGVTPLAVTL 311
Db 1034 NKAEKQKRDLGEELEA----LKTELEDTLDST-ATQQELRA------------------ 1069

Qy  312 DKGKLQLAPDNPPALNTLLKQTLGKDTQHYLAH------HASSDGSQHLLLDN----KG 360
Db 1070 -KREQEV-----------TVLKRALNEETRSHEAQVQEMRQKHAQAVQSLTEQLEQXKRAKA 1119

Qy  361 HLFDIKSTATSYSVLHNSHPGEIKGKLAQACTGSVSVDGKSGKISLGSGTQSHNKTMLSQ 420
Db 1120 NLDKNKQ?LEKENT-----------DLAGELRVLGQAKQEVEHRMKKLQ---AQVQELQSKCS 1168

Qy  421 PGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQAD 480
Db 1169 DGERARAEL------------NDKVHKLQNEV-------ESVTGMLNEAE 1199

Qy  481 GKLYAL-KDNRTLQN-LSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLD 538
Db 1200 GKAIKLAKDVASLSSQLQDTQE-----------LLQEESRQKLNVSTSLR 1238

Qy  539 ASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKL--FSAAIPKQGD 596
Db 1239 QLEEERNSLQ---DQLDEEMEAKQNLERH---ISTLNIQLSDSKKKLQDFASTVEALEE 1291

Qy  597 GNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPG 656
Db 1292 GKKRFPQKEI-------ENLTQQYEEKAAAYDKLEK----TKNRLQQE----------- 1327

Qy  657 WNLTDALV-IDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCK 715
Db 1328 --LDDLVVDLDNQRQL---------------VSNLEKKQRKFDQLLAEEKNISSKYA 1367

Qy  716 QLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNK-----PEPGDALQ 769
Db 1368 DERDRVEAEA--REKETKALSLARALEEALEAKEEL---ERTNKMLKAEMGRPGSA-- 1418

Qy  770 GLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIH 829
Db 1419 --SKDDVGQELSHDLEKSKRAL--GD-------PRLEEMKTQLEELGRTELASPRRDAK 1466

Qy  830 VDHKQNLYALTHEGEVFHQPREAWQNGAESSSWH-KLALPQSESKLKSLDMSHEHKPIAT 888
Db 1467 LRLEVNMQAPSRAS--FERDLQARTEQNEESRRHLQRQLHEYETELE----------- 1511

Qy  889 FEDGSQHQLKAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGVIPGSGLTVKLSAQ 948
```

```
Db 1512 -DERKQRALAAA------------AKIKLGWDPVRTLDLXADSAIKGR------- 1546
Qy  949 TGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQG 1008
Db 1547 -GGKAIKQLRKLQAQMKD----------FQRELEDARASRDEIFAT-----AKENEKKAKSLEA 1594
Qy 1009 ALIK-QLDAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDE---LEQSATRSVT 1064
Db 1595 DLMQLQEDLAAAEEGRKQADLEKE---ELAEELASSLSGRNALQDEKRRLEARIAQLEE 1650
Qy 1065 VLGQHQGVLKSNGEINSEFKPSPCKA-LVQSFNVNRSGQDLSKSLQQAVHATPPSAESK 1122
Db 1651 ELEEEQGNMEA--MSDRVRKATQQAEQLSNELATERSTAQKNESARQQLERQNKELRSK 1707
Qy 1123 LQSMLG---HFVSAGVDMSHQKGEIPLGRQRPDNDKTALTKSRLILDTVTIGELHELAD 1178
Db 1708 LHEMEGAVKSKFKSTIAALEAKIAQLEQVEQEAREKQAATKSLKQKDKKLKEILLQVED 1767
Qy 1179 KAKLVSDHKPDAD---QIKQLRQQFDTLRE-KRYESNPVKHYTDMGFTHNKALEANYD 1232
Db 1768 ERKMAEOYKEQAEKGNARVKQLKRQLEEAEEESQRINANRRKLQREL--DEATESN-E 1822
Qy 1233 AVKAFTNAFKKEHHGVNLTT--------RTVLESQGSAE 1263
Db 1823 AMGREVNALKSKLRRGNETSFVPSRRSGGRRVIENADGSE 1862
```

```
RESULT 5
PCT-US01-04098A-1662
; Sequence 1662, Application PC/TUS0104098A
; GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
;   FILE REFERENCE: 21272-029
;   CURRENT APPLICATION NUMBER: PCT/US01/04098A
;   CURRENT FILING DATE: 2001-02-05
;   PRIOR APPLICATION NUMBER: Not Yet Assigned
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: 09/728,422
;   PRIOR FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: 09/693,325
;   PRIOR FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 09/663,561
;   PRIOR FILING DATE: 2000-09-15
;   PRIOR APPLICATION NUMBER: 09/654,936
;   PRIOR FILING DATE: 2000-09-01
;   PRIOR APPLICATION NUMBER: 09/620,325
;   PRIOR FILING DATE: 2000-07-19
;   PRIOR APPLICATION NUMBER: 09/598,075
;   PRIOR FILING DATE: 2000-06-20
;   PRIOR APPLICATION NUMBER: 09/560,875
;   PRIOR FILING DATE: 2000-04-27
;   PRIOR APPLICATION NUMBER: 09/496,914
;   PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 1662
;   LENGTH: 2263
;   TYPE: PRT
;   ORGANISM: Homo sapiens
PCT-US01-04098A-1662

Query Match       2.1%;   Score 194;  DB 1;  Length 2263;
Best Local Similarity  19.5%;  Pred. No. 0.0066;
Matches 322;  Conservative 169;  Mismatches 563;  Indels 594;  Gaps 77;
PCT-US01-04098A-1662

Qy  147 RPMVKGGSGEDKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHH 206
Db  220 RPLKMGGR-KVEFPTDPKSRAQPPEDNRRKPVLGKLGTLFTAG--RRRNSRNGL---- 270
Qy  207 EIKPEPVGSTSKA------TTAHADRVEIAQEDDDSEFQQLHQQRLARE---RENPPQ-- 255
```

```
Qy   1122 KIQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLIIDTVTIGELHELADKAK 1181
                                                                         ------  1229
Db   1215 SPASLLN--ISAGSDDS--------------------------------------------

Qy   1182 LVSDHKPDADQIKQLRQQFDTL-----REKRYESNPVKHYTDMGFTHNKALEANYD-A 1233
               :|:||                    : :||:                :::    ||
Db   1230 -VFDSSSDMEKFTEIIKQMDSAVCMPMRKKARMPNSPAPHFA-MPPIHEDHLEKVFDPK 1287

Qy   1234 VKAFINAFKKE----------HHGVNLTTRTVLE-SQGSAELAKKLKNTLLSLDSGESMSF 1283
               ||   |||            ||:         ||  :: |:||   |:: :
Db   1288 VFTFGLGKKKESQPEMSPALHLMQNLDTKSKLRPKRASAEQSVLFK---------SL 1335

Qy   1284 SRSYGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGGVSGN 1343
                                                                    ------  TNGGIKRSRLEKSALFSS 1377
Db   1336 HTNTNGNSEPLVMPEINDKENRDV-

Qy   1344 IMVATGHDVM--PYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFK 1401
               :::|                                       :::||:::
Db   1378 LLSSLPQDKIFSPSVTSVNT-----------------------MTTAFSTSQNGSLSQSSVSQ 1417

Qy   1402 LTEDELPGFIHGLTHGTLTPAELLQ-----------KGIEHQMKQGSKLTFSVDTSA 1447
               :|| |                      : | ||  |::  :||
Db   1418 PTTEGAPPCGLNKEQSNLLPDNSLKVFNFNSSSTSHSSLKSPSHMEKYPQK---EKTKE 1473

Qy   1448 NLDLRAGINL----------------NEDGSKPNGVTARVSAGLSASAN----------  1480
               :|::|                    |||:|
Db   1474 DLDSRSNLHLPETKFSELSKLKNDDMEKANHIESVIKSNLPNCANSDTDFMGLFKSSRYD 1533

Qy   1481 --LAAGSRERSTTSGQFGSTTSASNNRP 1506
                 ::|:|:|  :|::::|| : ||
Db   1534 PSISFSGMSLSDTMTLRGSVQNKLNPRP 1561
```

RESULT   6
PCT-US01-04098A-3914
; Sequence 3914, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 3914
; LENGTH: 2515
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-04098A-3914

Query Match   2.0%;   Score 193.5; DB 1; Length 2515;
Best Local Similarity   17.9%; Pred. No. 0.0086;
Matches   377; Conservative 286; Mismatches 286; Indels 806; Gaps 94;
```
Db   271 ---ESPTRSNAKPLSPKDVVASPKLPERESERSRSQSSQLKQTDTSEGSPRENPREAEG 327
Qy   256 -PPKLGVATPISARFQPKLTAVAESV-LEGTDTTQSP-LKPQSMLKGSGAGVTPLAVTL 311
Db   328 ELPESGGPAAPPDAELSPRWSSSAAAVAVQQCHENDSPQLEP--LEAEGEPF-PDATTT 383
Qy   312 DKGKLQLAPDN----------PPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHLF 363
Db   384 AK-QLHSSPGNSSRQENAETPA------RSPGEDASPGAGH---EQEAFLGVRG---- 427
Qy   364 DIKSTATSYSVLHNSHPG------EIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTML 418
Db   428 ------------APGSPTQERPAGGLGEAPNAAPSVCAEEGSLGPRNARSQPPKGAS 472
Qy   419 SQPGE-------AH--RSLLTGIWQHPAGAARP-------QGESIRLHDDKIHIL 457
Db   473 DLPGEPPAEGAAHTASSAQADCTARPKGHAHPAKVLTLDIYLSKTEGAQV----DEPVVI 528
Qy   458 HP-----ELGVWQSADKDTHSQLSRQ----------ADGKL-------YALKDNRTLQN- 494
Db   529 TPRAEDCGDWDDMEKRSSGRRSGRRRGSQKSTDSPGADAELPESAARDAVFDDEVAPNA 588
Qy   495 LSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADA 554
Db   555 HQGLLHGKSE-------------LEAQSVAISHGRLV------------ 578
Qy   589 ASDNASAEK----KVKSPRAALDGGVASA------------ASPESKPS--------PG 623
Db   624 TKGQLRGESDRSKQPPASSPTKRKGRSRALEAVPAPPASGPRAPAKESPPKRVPDPSPV 683
Qy   579 -VADSEGKLFSAAIPKQGDNELKMKA-------MPQH--ALDEHF------GHDHQ 619
Db   684 TKGTAAESGEEAAARAIPR-----ELPVKSSSLLPEIKPEHKRGPLPNHFNGRAEGGRSRE 738
Qy   620 IS-------GFFHDD------------------HGQLNALVKNNFRQQHACPL 647
Db   739 LGRAAGAPGASDADGLKPRNHFGVGRSTVTTKVTLPAKPKHVELNLKTPKNLDS------L 793
Qy   648 GNDHQ------FHPGWNLTDALVIDNQLGLHHTNPEPHEILLDMGHLGSLA----LQEGKL 697
Db   794 GNEHNPFSQPVHKGNTATKISLFENK---RTNSSPRHTDIRGQRNTPASSKTFVGRAKL 849
Qy   698 HYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSLPH 757
Db   850 N----LAKKAKEMEQPEKKVMPNSPQNGVLVKETAIR-------TKVTVSEEI--LPA 895
Qy   758 VRNKPEPGDALQGL---NKDDKAQ------------AMAVIGVNKYILALTEKGDIRSFQ 801
Db   896 TRGMNGDSSENOALGPQPNQDKADVQTDAGCLSEPVASALIPVKDH-KLLEKEDSEAAD 954
Qy   802 IKPGTQQLERPAQTLSREGISGELKDI-HVDHKQNLYALTHEGEVFHQPREAW--QNG 856
Db   955 SK-------SLVLENVTDTAQDIPTTVDTKD--------LPPTAMPKPQHTFSDSQSP 997
Qy   857 AESSSWHKLAL--------------------PQSESKLKSLDMSHEHK----PIATFEDGSQHQ 896
Db   998 AESSPGSLSLSAPAPGDVPKDTCVQSPISSFPCTDLKVSENHKGCVLPVSRQNNEKMPL 1057
Qy   897 LKAGGWHAYAAP---ERGPLAVGTSGSQTVFNRL---------MQGVKGKVIPGSGLTV 943
Db  1058 LELGG---ETTPPLSTERSPEAVGSECPSRVLVQVRSFVLPVESTQDVSSQVIPES---- 1110
Qy   944 KLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPT--MSTPRPIKNAAYATQHGWQGREGLK 1001
Db  1111 ---------SEVREVQL------PITCHSNEPEVVSVASCA-----------P 1136
Qy  1002 PLYEMQGALIKQLDAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATR 1061
Db  1137 PQEEVLGNEHSHCTAELAAKSGPQ----------------------VIPPASEK 1168
Qy  1062 SVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAES 1121
Db  1169 TLPIQAQSQG---SRTPLMAESSPT---------NSPSSGNHL--ATPQRPDQTVTNGQD 1214
```

```
Qy   901  ------GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTQGMTG  954
Db  1210  DEGLIIGTHS---RNNPLHVGAEASE-                                  1246

Qy   955  AEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQL 1014
Db  1247  TEGFAESETFLTSTK----------------EGESGECAVAESEDRAADLL          1281

Qy  1015  DAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRD-ELEQSATRSVTVLGQHQGVL 1073
Db  1282  AVHAVKIEANVNSVTEEKDDAVTSAGSEEKCDGSLSRDSEIVEGTITFISEVESDGAVT  1341

Qy  1074  KSNGEI----NSEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSML   1127
Db  1342  SAGTEIRAGSISSEEVDGSQG-----NMMRMGP--KETEGTVTCTGAEGRS-         1386

Qy  1128  GHFVSAGVDMSHQKGE----IPLGRQRDPNDKTALTKSRLILDTVIGELHELADKA     1180
Db  1387  DNFVICSVTGAGPREERMVTGAGVVLGDNDAPPGTSASQEGDGSVNDGTEGE-SAVTST  1444

Qy  1181  KLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINA 1240
Db  1445  GITEDGEGPA-----------------SCTGSEDSSEGFAISSESENGES--AMDST    1483

Qy  1241  FKKEHHGVNLTT----RTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTV   1294
Db  1484  VAKEGTNVPLVAAGPCDDEGIVTSTGAKEEDEEGEDVVTSTGRGNEI---GHASTC     1536

Qy  1295  FVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGVSGNIMVATGHDVMP  1354
Db  1537  --------------------TGLGEESEGVLICESAEGDSQI----GTVEHVEAEAGAAIM- 1574

Qy  1355  YMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGL 1414
Db  1575  NANENNVDSMSGTEKGSKDTDICSSAKGIVESSVTSAVSGK---DEVTP-VPGG       1624

Qy  1415  THGTLTPAELLQKGIEHQMKQGSKLTFSVDTSANLDLRAG-INLNEDGSKPNGTARVSA  1473
Db  1625  CEGPMTSAASDQ------------SDSQLEKVEDTTISTGLVGGSYDVLVSGEVPECEVAHTSP 1676

Qy  1474  GLSASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGK 1533
Db  1677  SEKEDEDIIT-SVENEECDGLMATTASGD----ITNQNSLAGGKNQGKVLIISTSTTNDYT 1732

Qy  1534  PVGIFPAFTSTNVSAALALDNRTSQSI------SLELKRAEP-VTSNDISELT-        1579
Db  1733  P----QVSAITDVEGGLSDALRTEENMEGTRVTTEEFEAPMPSAVSGDDSQLTASRSEEK 1788

Qy  1580  --------STLGKHFK---DSATT------KMLAALKE-                      1605
Db  1789  DECAMISTSIGEEFELPISSATTIKCAESLQPVAAAVEERATGPVLISTADFEGPMPSAP 1848

Qy  1606  PAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLSR 1665
Db  1849  PEAESPLASTSKEEKDECA------------LISTSIAEECEASVSGVVVESENERAGTVMEE 1899

Qy  1666  INNDGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPF---SSASVSME  1721
Db  1900  KDGSGIISTSSVE-DCEGPVSSAVPQEE---GDPSV-----TPAAEMGDTAMISTS     1946

Qy  1722  LKDGLRREQTEKAILDGKVGREEVGVLFQDRNNL---RVKSVSVSQSVSKSEGFNTPALL 1778
Db  1947  TSEGC---EAVM------IGAVLQDEDRLTITRVEDLS-                      1980

Qy  1779  GTSNSAAMSMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASAL-TDLKKEGLEM  1836
Db  1981  STSTAECMPISASI----------DRHEENQLTAD------NPEGNGDLSATEVSKHKVPM 2025

Qy  1837  KS 1838
Db  2026  PS 2027
```

```
Qy    28  LQQGSSSSPQNAAASLAA------EGKNRGKMPRIHQPSTAADGI-SAAHQQKKSFSLR   80
Db   268  LKHVHAKSEPSKPARRLSESLHVVDENKNESKIEREHKRRTSPVIMEGVQEETDTRDVK  327

Qy    81  GCLGTKKFSRSAPQGQPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLTRSGGVKRRN 140
Db   328  RQVERSEICTEPQKQSTLKNE-------KHLKKDDSETPHLKS--------LLKKE     370

Qy   141  MDDMAGRPMVKGGSGEDKVPTQQK------RHQLNNFGQMR------QTMLSKMAHPAS  187
Db   371  VKSSKEKPEREKTPSEDKLSVKHKYKGDCMHKTGDETELHSSEKGLKVEENIQKSQQTK  430

Qy   188  ANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRV-EIAQEDDSEFQQL-----  241
Db   431  LSSDDKTERKSKH------RNERKLSVLGKDGKPVSEYIIKTDENVRKENNKKERRLSAEKT 486

Qy   242  --HQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLE------GTDTT--QSPLKP 293
Db   487  KAEHKSR--RSSDSKIQKDSLG-----SKQHGITLQRRSESYSEDKCDMDSTNMDSNLKP 539

Qy   294  QSML-------------KGSGAGVTPLAVTLDKGKLQLAPDNPP------AL 326
Db   540  EEVVHKEKRRTKSLLEEKLVLKSKSKTQGKQVKVVETELQEGATKQATTPKDKENTEE  599

Qy   327  NTLLKQTLGK--------DTQHYLAHHASSDGSQH--LLLDNKGHLFDIKSTA 369
Db   600  NDSEKQRKSKVEDKPFEETGVEPVLETASSSAHSTQKDSSHRAKLPLAKEKYKSDKDSTS 659

Qy   370  T--------------SYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNK-- 415
Db   660  TRLERKLSDGHKSRSLKHSSK--DIKKK-------DENKSDDKDGK---EVDSSHEKARG  707

Qy   416  --------TMLSQPGEAHRSLLTGIWQHPAGAA--RPQ--GESTRLHDD 452
Db   708  NSSLMEKKLSRRLCENRRGSLSQEMAKGEEKLAANTLSTPSGSSLQRPFKKSGDMTLIPEQ 767

Qy   453  KIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYS  512
Db   768  EPMEIDSEPGV-------ENVFEVSKTQDNRNNNSHQDIDSENMKQKTSATVQKDELRTCT 821

Qy   513  VDQRGQVAILTDTPGRHKMSIMPSLDASPESH------ISLSLHFADA--------HQGL  558
Db   822  ADSKATAPAY--KPGRG-------TGVNSNSEKHADHRSTLTKMHIQSAVSKMNPGEKEPI 874

Qy   559  LHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGH--  616
Db   875  HRGTTEVNIDSETV-HRMLLSAPSE-------------NDRVQOKNLKNTAAEEHVAQGD  919

Qy   617  ---DHQISGFFHDDHGQLNALVKNNFRQQH---ACPLGN---------DHQFHP 655
Db   920  ATLEHSTN---LDSSPSLSSVTVVPLRESYDPDVIPLFDKRTVLEGSTASTSPADHSALP 976

Qy   656  GWNLIT---------DALVIDNQLGLHHTN---PEPHE--ILDMGHIGSLALQ 693
Db   977  NQSLTVRESEVLKTSDSKEGGEGFTVDTPAKASITSKRHIPEAHQATLLD-GKQGKVIMP 1035

Qy   694  EG-KLHYFDQLTKGWTGAESDCKOLKKGLDGAAYLLKDGEVKRLNINQSTS--SIKHGT  749
Db  1036  LGSKL-------TGVIVENENITK--EGG--LVDMAKKENDLNAEPNLKQTIKATV    1080

Qy   750  ENVFSLPHYVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQL 809
Db  1081  EN------------GKKDGIAVDHVGLN-                               1098

Qy   810  ERPAQT--LSREGISGELKDIHVD---------HKQNLYALTHEG----EVFHQ 848
Db  1099  EKYAETVKLKHKRSPGKVKDISIDVERRNENSEVDTSAGSGSAPSVLHQRNGQTEDVATG 1158

Qy   849  PREAWQNGAESSSWHKLALPLPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG------ 900
Db  1159  PRRAEKTSVATS-------TEGKDKDVTLSPVKAGPATTTSSETRQSEVALPCTSIEA  1209
```

```
RESULT 7
PCT-US01-04098A-3915
;  Sequence 3915, Application PC/TUS0104098A
;  GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
;   FILE REFERENCE: 21272-029
;   CURRENT APPLICATION NUMBER: PCT/US01/04098A
;   CURRENT FILING DATE: 2001-02-05
;   PRIOR APPLICATION NUMBER: Not Yet Assigned
;   PRIOR APPLICATION FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: 09/728,422
;   PRIOR APPLICATION FILING DATE: 2000-11-30
;   PRIOR APPLICATION NUMBER: 09/693,325
;   PRIOR APPLICATION FILING DATE: 2000-10-20
;   PRIOR APPLICATION NUMBER: 09/663,561
;   PRIOR APPLICATION FILING DATE: 2000-09-15
;   PRIOR APPLICATION NUMBER: 09/654,936
;   PRIOR APPLICATION FILING DATE: 2000-09-01
;   PRIOR APPLICATION NUMBER: 09/620,325
;   PRIOR APPLICATION FILING DATE: 2000-07-19
;   PRIOR APPLICATION NUMBER: 09/598,075
;   PRIOR APPLICATION FILING DATE: 2000-06-20
;   PRIOR APPLICATION NUMBER: 09/560,875
;   PRIOR APPLICATION FILING DATE: 2000-04-27
;   PRIOR APPLICATION NUMBER: 09/496,914
;   PRIOR APPLICATION FILING DATE: 2000-02-03
;   NUMBER OF SEQ ID NOS: 3960
;   SOFTWARE: Custom
;   SEQ ID NO 3915
;   LENGTH: 2515
;   TYPE: PRT
;   ORGANISM: Homo sapiens
PCT-US01-04098A-3915

Query Match       2.0%;    Score 193.5;  DB 1;  Length 2515;
Best Local Similarity  17.9%;   Pred. No. 0.0086;
Matches  377;  Conservative 286;  Mismatches 806;  Indels 633;  Gaps 94;

Qy    28 LQQGSSSSSPQNAAASLAA------EGKNRGKMPRIHQPSTAADGI-SAAHQQKKSFSLR 80
             |:   ||  |||  |:|         |||  :| |    :          |:
Db   268 LKHVHAKSEPSKPAARRLSESLHVVDENKNESKIEREHKRRTSTPVIMEGVQETDTRDVK 327

Qy    81 GCLGTYKKFSRSRAPQGQPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLTRSGGVKRRN 140
               : ||  :||            ||| |   |:       ||   :: ||:
Db   328 RQVERSEICTEEPQKQKSTLKNE----KHLKKDDSETPHLKS-------LLKKE 370

Qy   141 MDDMAGRPMVKGGSGEDKVPTQQK---RHOLNNFGQMR--------QTMLSKMAHPAS 187
             |:|| |  : ::  | ||           |  :            | |  |
Db   371 VKSSKEKPEREKTPSEDKLSVKHKYKGDCMHKTGDETFLHSSEKGLKVEENIQKQSQQTK 430

Qy   188 ANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKATTAHADRV-ELAQEDDDSEFQQL----- 241
              :       |   |        |   | : |  :   :: || |: :
Db   431 RNERKLSVLGKDGKPVSEYIIKTDENVRKENNKKERRLSAEKT 486

Qy   242 --HQQRLARERENPPQPPKLGVATPISARFQPKLTAVAESVLE--GTDTT--QSPLKP 293
            |   | :|      |   |||    | |:          :   |||:|:  ||||
Db   487 KAEHKSR--RSSDSKIQKDSLG------SKQHGITLQRRSESYSEDKCDMDSTNMDSNLKP 539

Qy   294 QSML----------------------------KGSGAGVTPLAVTLDKGKGLQLAPDNPP-----AL 326
              : :                          |  :| ||         | |   |
Db   540 EEVVHKEKRRTKSLLEEKLVLKSKSKTQGKQVKVVETELQEGATKQATTPKPDKEKNTEE 599

Qy   327 NTLLKQTLGK------DTQHYLAHHASSDGSQH-LLLDNKGHLFDIKSTA 369
             :| || |:       |||:  ||  | |||  | |   |
Db   600 NDSEKQRKSKVEDKPFEETGVEPVLETASSSAHSTQKDSSHRAKLPLAKEKYKSDKDSTS 659

Qy   370 T-------SYSVLHNSHPGEIKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNK--- 415
                     :| |   :|:    |  :    | |:   || :   |:
Db   660 TRLERKLSDGHKSRSLKHSSK--DIKKK------DENKSDDKDGK-----EVDSSHEKARG 707

Qy   416 --------TMLSQPGEAHRSLLTGIWQHPAGAA--RPQ--GESIRLHDD 452
                    |   | :|  | |    | |   |     |||   |  |:|:
Db   708 NSSLMEKKLSRRLCENRRGSLSQEMAKGEEKLAANTLSTPSGSSLQRPKKSGDMTLIPEQ 767

Qy   453 KIHILHPELGVWQSADKDTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYS 512
              :    |::|   |   :  |  |    :| |  | :    | :  |:| |
Db   768 EPMEIDSEPGV------ENVFEVSKTQDNRNNNSHQDIDSENMKQKTSATVQKDELRTCT 821

Qy   513 VDQRGQVAILDTPGRHKMSIMPSLDASPESH------ISLSLHFADA--------HQGL 558
             | :    :|  |  |  |     |      |      :|  |
Db   822 ADSKATAPAY--KPGRG-----TGVNSNEKHADHRSTLTKKMHIQSAVSKMNPGEKEPI 874

Qy   559 LHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQHALDEHFGH-- 616
             |   |    | |:        |  :|       | |  ||  |:
Db   875 HRGTTEVNIDSETV-HRMLLSAPSE------NDRVQKNLKNTAAEEHVAQGD 919

Qy   617 ---DHQISGFFHDDHGQLNALVKNNFRQQH--ACPLGN------DHQFHP 655
                |   | :   | |  |    |    :   | |:|       ||
Db   920 ATLEHSTN---LDSSPSLSSVTVVPLRESYDPDVIPLFDKRTVLEGSTASTSPADHSALP 976

Qy   656 GWNLT-----DALVIDNQLGLHHTN----PEPHE--ILDMGHLGSLALQ 693
                |:      |  :|:  | |       | |:   :|  | |
Db   977 NQSLTVRESEVLKTSDSKEGEGEGFTVDTPAKASITSKRHIPEAHQATLLD-GKQGKVIMP 1035

Qy   694 EG-KLHYFDQLTKGWTGAESDCKQLKGLDGAAYLLKDGEVKRLNINQSTS--SIKHGT 749
             :|  |    | |       |   |  :|:          |  | | :|    ||||
Db  1036 LGSKL-----TGVIVENENITK--EGG--LVDMAKKENDLNAEPNLKQTIKATV 1080

Qy   750 ENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSFQIKPGTQQL 809
             ||    |   |::    | | |:          || |  |||
Db  1081 EN-----GKKDGIAVDHVGLN------------------T 1098

Qy   810 ERPAQT-LSREGISGELKDIHVD---HKQNLYALTHEG----EVFHQ 848
             |   |: |   |:  | |:  |    || | |||  |     |:|
Db  1099 EKYAETVKLKHKRSPGKVKDISIDVERRNENSEVDTSAGSGSAPSVLHQRNGQTEDVATG 1158

Qy   849 PREAWQNGAESSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAG-- 900
             |   :|        | |    | |    :    |  |||
Db  1159 PRRAEKTSVATS------TEGKDKDVTLSPVKAGPATTSSETRQSEVALPCTSIEA 1209

Qy   901 ------GWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLSAQTGGWTG 954
                   |::  |||  | | ||||    |  |      |   ||:||  ||  ||| ||
Db  1210 DEGLIIGTHS---RNNPLHVGAEASE------CTVFAAAEEGGAVV 1246

Qy   955 AEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQL 1014
             :|| :| |:|  |||||||| |||||||:||   |: |:|  |
Db  1247 TEGFAESETFLTSTK--------EGESGECAVAESEDRAADLL 1281

Qy  1015 DAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRD-ELEQSATRSVTVLGQHQGVL 1073
             :||:|::||  |  :  ||   :   |   |:  |:||   ||:|  |
Db  1282 AVHAVKIEANVNSVVTEKDDAVTSAGSEEKCDGSLSRDSEIVEGTITFISEVESDGAVT 1341

Qy  1074 KSNGEI------NSEFKPSPGKALVQSFNVRSGQDLSKSLQQAVHATPPSAESKLQSML 1127
             |||:|       |  | ||| |  | :| |||||||||| ||  ||||| | |||||
Db  1342 SAGTEIRAGSISSEEVDGSQG------NMMRMGP--KKETEGTVTCTGAEGRS------ 1386

Qy  1128 GHFVSAGVDMSHQKGE-------IPLGRQRDPNDKTALTKSRLILDTVTIGELHELADKA 1180
             |:|:|:|  | |:         ||||| ||||||||||| ||||| |:|||  |||||
Db  1387 DNFVICSVTGAGPREERMVTGAGVVLGLDNDAPPGTSASQEBDGSVNDGTEGE--SAVTST 1444

Qy  1181 KLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINA 1240
             | |||||||| |:|  |   ||||| ||||| |  || |:||||||||  || |:|
Db  1445 GITEDGEGPA------------SCTGSDEDSSEGFAISSESEENGES--AMDST 1483

Qy  1241 FKKEHHGVNLTT-------RTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTV 1294
             |||::|| |||        | | ||| ||||||||||||||| | |   |  |   ||:
Db  1484 VAKEGTNVPLVAAGPCDDEGIVTSTGAKEEDEEGEDVVTSTGRGNEI------GHASTC 1536

Qy  1295 FVPTLSKKVPVPVIPGAGITLDRAYNLSFRSTSGGLNVSFGRDGGVSGNIMVATGHDVMP 1354
             |   |||| || |||||| || |||| | ||||| | |  |  | |||||| ||| | |
Db  1537 -----------------TGLGEESEGVLICESAEGDSQI----GTVVEHVEAEAGAIM- 1574

Qy  1355 YMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLFKFLTEDELPGFIHGL 1414
```

```
Db    1575 ------NANENNVDSMSGTEKGSKDTDICSSAKGIVESSVTSAVSGK---DEVTP-VPGG 1624

Qy    1415 THGTLTPAELLQKGIEHQMKQGSKLTFSVDTSANLDLRAG-INLNEDGSKPNGVTARVSA 1473
           :| : || | | ||  | |||    :|:   :|:   |:| | |||| :||  || | |
Db    1625 CEGPMTSAASDQ------SDSQLEKVEDTTISTGLVGGSYDVLVSGEVPECEVAHTSP 1676

Qy    1474 GLSASANLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHEGK 1533
           || |  || :||   |  :   ||   |:|| | || | |  | |  |: || || |:
Db    1677 SEKEDEDIIT-SVENEECDGLMATTASGD--ITNQNSLAGGKNQGKVLIISTSTTNDYT 1732

Qy    1534 PVGIFPAFTSTNVSAALALDNRTSQSI------SLELKRAEP-VTSNDISELT------ 1579
           || |  | || |||  |  ||| :|:|       :|: |:|| |||| :|:|
Db    1733 P----QVSAITDVEGLSLSDALRTEENMEGTRVTTEEFEAPMPSAVSGDDSQLTASRSEEK 1788

Qy    1580 ------STLGKHFK---DSATT------KMLAALKE-------------------LDDAK 1605
              |:|| | |  :|  ||:|| ||         |:| ||  ::||
Db    1789 DECAMISTSIGEEFELPISSATTIKCAESLQPVAAAVEERATGPVLISTADFEGPMPSAP 1848

Qy    1606 PAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQOAADSHSMELGSASHSTTYNNLSR 1665
           :   |   :| |          | || | :     ||     :     || |:  : :
Db    1849 PEAESPLASTSKEEKDECA-------LISTSIAEECEASVSGVVVESENERAGTVMEE 1899

Qy    1666 INNDGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLOSTPF----SSASVSME 1721
           |  | ||| || | |  |:| | |  | | |::  | |:|:|  |:|    ::|:|:|
Db    1900 KDGSGIISTSSVE-DCEGPVSSAVPQEE---GDPSV-------TPAEEMGDTAMISTS 1946

Qy    1722 LKDGLREQTEKAILDGKVGREEVGVLFQDRNNL----RVKSVSVSQSVSKSEGFNTPALLL 1778
           | :| :|  ||: |:|| ||| :| :: :        |: : :|| : |: |::|||||
Db    1947 TSEGC-----EAVM-------IGAVLQDEDRLTITRVEDLS----------DAAII 1980

Qy    1779 GTSNSAAMSMERNIGTINFKYGQDQNTPRRFTLEGGIAQANPQVASAL--TDLKKEGLEM 1836
           ||| | |  |       |:|     | | | | | |    ||| | |  | |: :|:|
Db    1981 STSTAECMPISASI---------DRHEENQLTAD------NPEGNGDLSATEVSKHKVPM 2025

Qy    1837 KS 1838
           ||
Db    2026 PS 2027
```

RESULT 8
PCT-US01-03782A-284
; Sequence 284, Application PC/TUS0103782A
; GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   APPLICANT: Ford, John E et al
;   TITLE OF INVENTION: Novel Bone Marrow Nucleic Acids and Polypeptides
;   FILE REFERENCE: 21272-040
;   CURRENT APPLICATION NUMBER: PCT/US01/03782A
;   CURRENT FILING DATE: 2001-02-05
;   PRIOR APPLICATION NUMBER: 09/496,914
;   PRIOR FILING DATE: 2000-02-03
;   PRIOR APPLICATION NUMBER: 09/598,075
;   PRIOR FILING DATE: 2000-06-20
;   PRIOR APPLICATION NUMBER: 09/620,325
;   PRIOR FILING DATE: 2000-07-19
;   PRIOR APPLICATION NUMBER: 60/250,583
;   PRIOR FILING DATE: 2000-11-30
;   NUMBER OF SEQ ID NOS: 386
;   SOFTWARE: Custom
; SEQ ID NO 284
;   LENGTH: 5447
;   TYPE: PRT
;   ORGANISM: Homo sapiens
PCT-US01-03782A-284

Query Match   2.0%; Score 190; DB 1; Length 5447;
Best Local Similarity 17.6%; Pred. No. 0.058;
Matches 415; Conservative 311; Mismatches 806; Indels 820; Gaps 111;

```
Qy    27 ALQQGSS-----SSSPQNAAASLAA------EG--KNRGKM-PRIHQPSTAADGISAAHQ 72
          || ||||          :| :  | |     ||  | | :  | :| ||  || :| |
Db  1601 ALQKNQSDLKDLQDDIQNRATSFATVVKDIEGFMEENQTKLSPR--ELTALREKLHQAKE 1658
```

```
Db    73 OKKSFSLRGCLGTKKFSRSAPQG-QPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLT 131
          :|:| || |         :||    : | : :| | |||:|      |   |   | ||
Qy  1659 QYEALQEETRVAQKELEEAVTSALQQETEKSKAA--KELA-ENKKKIDALLDWVTSVG 1713

Db   132 RSGGVKRRN--MDDMAGRPMVKGG--------------SGEDKVPTQ------QL 168
          | :|  ||:   ||:|| :|  | |              || | |  |       ::|
Qy  1714 SSGGQLLTNLPGMEQLSGASLEKGALDTTDGYMGVNQAPEKLDKQCEMMKARHQELLSQQ 1773

Db   169 NNF--GQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKA-TTAHA 224
          ||:    | | |     :: |    | :     |       :|:|  ||:||  |:||
Qy  1774 QNFILATQSAQAFLDQHGHNLTPEEQOMLQQ------KLGELKEQ-YSTSLAQSEAEL 1824

Db   225 DRVEIAQEDDSEFQQLHQ---------QRLARERENPPQPPKLGVATPISARFQPKLTAVAE 278
          | :| :: ::| | | |          :| |:||| | | | |  | :  ||  |  ||:
Qy  1825 KQVQTLQ-DELQKFLQDHKEPESWLERSEKELENMHK----GGSSP-------- 1865

Db   279 SVLRGTDTTQSPLKPQSML--------------KG-------SGAGVTPLAVTLDKGKLQLAPDNPP 324
          | |   || :|| |   ::              ||        : | |    | :|| | |:
Qy  1866 ------ETLPSLLKRQGSFSEDVISHKGDLRFVTISGQKVLDMENSPKEGK-----EPS 1913

Db   325 ALNTLLKQTLGKDTQHYLAHHASSD--GSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGE 382
          |  | || | || | |  | || |   | :|   ||  ||  |: |::|| |  :| |
Qy  1914 EIGNLVKDKLKDATERYTALHSKCTRLGSHLNMLLGQYHQF--QNSADSLQAWMQACEAN 1971

Db   383 IKGKLAQAGTGSVSVDGKSGKISLSGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARP 442
          |:|| | | | :|||  ||| || : |:| | | :|  |   :||  |    |  |
Qy  1972 VEKLLSD-----------------TAASDPGVLQEQLAT-----------TKQ 1996

Db   443 QGESIRLHD--DKIHILHPELGVWQSADKDTHSQLSRQADGKL-------YALKDNRTL 492
          |||| | |   |:| |  | | |  |  ||  |  |:|| | |       | |:| |
Qy  1997 LQEELAEHQVPVEKLQKVARDIMEIEGEPAPDHRHVQETTDSILSHFQSLSYSLAERSSL 2056

Db   493 ------QNLSDNKSSEKLVDKIKSYSVDQRG-QVAILTDTPGRHKMSIMPSLDASPESHI 545
                 |: | : |::|  |  : |||  |  |||  || |  |:  ||| ||: |::
Qy  2057 LQKAIAQSQSVQDSLESLLQSIGEVEQNLEGKOVSSLSSGVIQEALATNMKL-------- 2108

Db   546 SLSLHFADAHQGLLHGKSELEA-------------QSVAISHGRLVVADSEGKLFSAAI 591
          :| | | : | |  | |::|                |  :|| || | |: ||: :|:
Qy  2109 -------------KQDIARQKSSLEATREMVTRFMETADSTTRAVLQGKL--AEVSQRFEQLCL 2157

Db   592 PKQGDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDH 651
          | |  :|| |  |  : |  | | |:    ||  | | |  | |  | | | | |   |
Qy  2158 QQQEKESSLK-KLLPQAEMFEHLS--------------GKLQQFMENKSRM----LASGN 2198

Db   652 QFHPGWNLTDALVIDNQLGLHHTNPEPH----------EILDMGHGLSLALQEGKLHY-- 699
          || ||  | |  | |:| |  :|              | | || | | ||  :| ::
Qy  2199 Q--PDQDITHFQQIQELNLEMEDQQENLDTLEHLVTELSSCGFALDLCQHQDRVQNLRK 2256

Db   700 -FDQLTKGWTGAESDCKQLKKGLDGAAYLLK---------DGEVKRLNINQSTSSIKHGT 749
           | || | | |   |  ||| |||   |||          | | | ||:|:|::| |
Qy  2257 DFTELQKTVKEREKDASSCQEQLDEFRKLVRTFQKWLKETEGSIPPTETSMSAKELEKQI 2316

Db   750 ENVFSL-----PHVRNKPEPGDALQGL------NKDDKAQAMAVI----GVNKY 788
          || | |      | | | |  | | |       |:|| | ||       || |
Qy  2317 EHLKSLLDDWASKGTLVEEINYKGTSLENLIMEITAPDSQGKTGSILPSVGSSVGSVNGY 2376

Db   789 LA---LTE---KGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALTH 841
          ||   | |   | |   |  | | :|| |  |  |  |:  |    | |  :|  :| |
Qy  2377 HTCKDLTEIQCDMSDVNLKYEKLGGVLHERQE--SLQAILNRMEEVIKEANSVLQWLES 2433

Db   842 EGEV----------FHQPREA--------WQNGAE 858
          | :|           | |            | ||
Qy  2434 KEEVLKSMDAMSSPTKETVKAQAESNKAFLAELEQNSPKIQVKEALAGLLVTYPNSQE 2493

Db   859 SSSWHKL-----ALPQSESKLKSLDMSHEHKPIATFEDGSQHLKAGGWHAYAAP 908
          || || :       | ::| |  | |::| :   |  | | : | ||| ||
Qy  2494 AENWKKIQEELNSRWERATEVTVARQRLEESASH-LACFQ-AAESOLQP-W----LM 2544

Db   909 ER------GPLAV----GTSGSQTVF----------NRLMQGVGKVIPGSG--- 940
          |        |         | ||          |: : |: || ||
Qy  2545 EKELMMGVLGPLSIDPNMLNAQKQQVQFMLKEFEARRQQHEQLNEAAQG-ILTGPGDVSL 2603
```

```
Db  3598  IKT--RYADITVTSSKALRTLEQ------ARQLATKFQST  3641
Qy  1731  EKAILDGKVGREEVG--VLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLGTSNSAAMS  1787
Db  3642  EELATSG--GQSPTGEQIPQFQQRQKELKKEVMEHRLVLDTVNEVSRALLELVPWRAREG  3699
Qy  1788  MERNIGTINFKY  1799
Db  3700  LDKLVSDANEQY  3711
```

```
RESULT  9
US-09-196-296B-2
; Sequence 2, Application US/09196296B
; GENERAL INFORMATION:
;   APPLICANT: Suen, Chen-Shian
;   APPLICANT: Frail, Donald E.
;   APPLICANT: Lyttle, Richard C.
;   TITLE OF INVENTION: Cloning and Expression of a Nuclear
;   TITLE OF INVENTION: Receptor Coactivator Proteins and Uses Thereof
;   FILE REFERENCE: 0630/01375
;   CURRENT APPLICATION NUMBER: US/09/196,296B
;   CURRENT FILING DATE: 1998-11-19
;   NUMBER OF SEQ ID NOS: 3
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1423
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-196-296B-2

Query Match          2.0%;  Score 188.5;  DB 5;  Length 1423;
Best Local Similarity  17.8%;  Pred. No. 0.0064;
Matches 239;  Conservative 190;  Mismatches 421;  Indels 493;  Gaps 62;
```

```
Qy   681  ILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLL-KDGEVKRLNIN  739
Db   107  VIDKDSLGPLLLQ----------------------ALDGFLFVNRDGNIVFVSEN  140

Qy   740  QSTSSIKHGTE------NVFSLPH----------VRNKPE-  763
Db   141  -VTQYLQYKQEDLVNTSVYNIHEEDRKDFLKNLPKSTVNGVSWTNETQRQKSHTFNCRM  199

Qy   764  -------PGDALQGLNKD---DKAQAMAVIGNKYLALTEKG-DIRSF------QIKPGTQQ  808
Db   200  LMKTPHDILEDINASPEMRQRYETMQCFALSQPRAMMEEGEDLQSCMICVARRITTGERT  259

Qy   809  L-ERPAQTLSREGISGELKDIHVDHK--------------QNLYALTHEGEVFHQPR-  850
Db   260  FPSNPESFITRHDLSGKVVNIDTNSLRSSMRPGFEDIIRRCIQRFFSL-NDGQSWSQKRH  318

Qy   851  --EAWQNGAESSWHKLAL------------PQSESKLKSLDMSHEHKPIATFEDG--SQH--QL  897
Db   319  YQEAYLNGHAETPVYRFSLADGTIVTAQTKSKL----FRNPVTNDRHGFVSTHFLQR  371

Qy   898  KAGGWHAYAAPERGPLAVGTSGSQTVFNRLMQGVKGKVIPGSGLTVKLS---------AQT  949
Db   372  EQNGYR-----PNPNVPVQGIRPPMAGCNSSVGGM--SMSPNQGLQMPSSRAYGLADPSTT  425

Qy   950  GGMTGAE--GRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGW--QGREGLKPLYE  1005
Db   426  GQMSGARYGG-------SSNIASLTGPGMQSPSYQNNNYGLNMSSPPHGSPGLAP--N  476

Qy  1006  MQGALIKQLDAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTV  1065
Db   477  QQNIMISP-------RNRGSPKIASH--------  500

Qy  1066  LGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQS  1125
Db   501  AGVHSPMASSGNTGNHSFSSSSLSAL-------QAISEGVGTSLLLSTLSSPGPKLDN  550
```

```
Qy   941  -------------LTVKLSAQTGGMTGA------EGRKVSSKFSERIRAYAFNPTM  977
Db  2604  STSQVQKELQSINQKWVELTDKLNSRSSQIDQAIVKSTQYQELLQDLSEKVRAVG------  2658

Qy   978  STPRPIKNAAYATQHGWQGREGLKPLYEMQGAL--IKQLDAHNVRHNAP---------  1024
Db  2659  --QRLSVQSAISTQ---PEAVKQQLEETSEIRSDLEQLD-HEVKEAQTLCDELSVLIG  2710

Qy  1025  ----QPDLQSKLETLDLGEHGAE------LLNDMKRFRDELEQS  1058
Db  2711  EQYLKDELKRKRLETVALPLQGLEDLAADRINRLQAALASTQQFQQMFDELRTWLDDKQSQ  2770

Qy  1059  ATRSVTV---LGQHQG---VLKSNGE-INSEFKPSPGKALVQS  1094
Db  2771  QAKNCPISAKLERLQSOLQENEEFQKSLNQHSGSYEVIVAEGESLLLSVPPGEEKRTLQN  2830

Qy  1095  -FNVNRSGQDLSKS---------LQQA-------VHATP  1116
Db  2831  QLVELKNHWEELSKKTADRQSRLKDCMQKAQKYQWHVEDLVPWIEDCKAKMSELRVTLDP  2890

Qy  1117  PSAESKL--QSMLG----HFVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLI  1164
Db  2891  VQLESSLLRSKAMLNEVEKRRSLLETLNSSAADILINSSEADEDGIRD--EKAGINQN---  2945

Qy  1165  LDTVT--IGELHELADKAK----  1181
Db  2946  MDAVTEELQAKTGSLEEMTQRLREFQESFKNIEKKVEGAKHQLEIFDALGSQACSNKNLE  3005

Qy  1182  -------LVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHY  1216
Db  3006  KLRAQQEVLQALEPQVDYLRNFTQGLVED-APDGSDASQLLHQAEVAQQFLE---VKQR  3061

Qy  1217  TDMG--FTHNK------ALEANYDAVKAFI  1238
Db  3062  VNSGCVMMENKLEGIGQFHCRVREMFSQLADLDELDGMGAIGRDTDSLQSQIEDVRLFL  3121

Qy  1239  N---AFKEHHGVNLTTRTVLESQGSAEL-------AKKLKNTLLSLD  1276
Db  3122  NKIHVLKLDIEASEAECRHMLEEGTLDLLGLKRELEALNKQCGKLTERGKARQELEL  3181

Qy  1277  SGESMSFSRSYGG----------GVSTVFV-------PTLSKKVPVVIPGA  1311
Db  3182  LGRVEDFYRKLKGLNDATTAAEEAEALQWVGTEVEIINQQLADFKMFQKEQVDPL----  3237

Qy  1312  GITLDRAYNLSFSRTSG---GLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASD  1368
Db  3238  ------QMKLQQVNGLGQLIQSAGKDCDVQG-----LEHDMEEI--------NAR-  3274

Qy  1369  WLSAKHKIS---PDLRIGAAVSGTLQGTLQNSLKFKLTEDEL-----PGFIHGLTHGTL  1419
Db  3275  WNTLNKKVAQRIAQLQEALLHCGKFQDALEPLLSWLADTEELIANQKPPSAEYKVVKAQI  3334

Qy  1420  TPAELLQKGIEHQMKQGSKLTFSVDTSANLDRAGINLNEDGSKPNGVTARVSAGLSASA  1479
Db  3335  QEQKLLQRLLD--------DRKATVDM-----LQAEGGR-------IAQSA  3365

Qy  1480  NLAAGSRERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSSTHE-GKPVGIF  1538
Db  3366  ELA--DREKIT----GQLESLESRWTELLSKAAARQKLEDILVLAKQFHETAEPISDF  3418

Qy  1539  PAFTSTNVSAALALDNRTS--QSISLELKRAEPVTSN-------DISELTSTLG  1583
Db  3419  LSVTEKKLLANSEPVGTQTAKIQQQIIRHKALEEDIENHATDVHQAVKIGQSLSSLTSPAE  3478

Qy  1584  KHFKDSATTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGDERYEA---------  1630
Db  3479  QGVLSEKIDSLQARYSEIQD-RCCRKAALLDQALSNARLFGEDEVEVLNWLAEVEDKLSS  3537

Qy  1631  --VRNLKKLVIRQQAAD----SHSMELGSA-SHSTTYNNLSRINN-DG  1670
Db  3538  VFVKDFKQDVLHRQHADHLALNEEIVNRKKNVDQAIKNGQALLKQTTGEVLLIQEKLDG  3597

Qy  1671  IVELLHKHFDAALPASSAKRLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQT  1730
```

```
Qy   1126  MLGHFVSAGVDMSHOKGEIPLGRQRDPNDKTALTKSRLLDTVTIGELHELADKAKLVSD  1185
                 : :||          :||  ||   ::||| :| :|:
Db    551  SPNMNITQPSKVSNQDSKSPLGFYCDQN----------------------------------   578

Qy   1186  HKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEH  1245
                :|         : |         :: || ||
Db    579  -----------------PVESSMCQSNSRDHLSDK-------ESKESSVEGAEN-----------   608

Qy   1246  HGVNLTTRTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYGGVSTVFVPTLSKKVPV  1305
                 :||:|           ||   :|:|::|
Db    609  --QRGPLESKGHKKLLQLL--TCSSDRGHS-----------SLTNSPLDSSCKESSV  651

Qy   1306  PVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGGV----------------SGN----I  1344
                 :|::            |||||::|          ||          :||
Db    652  SVTSPSGVS----------SSTSGGVSSTSNMHGSLLQEKHRILHKLLQNGNSPAEVAKIT  702

Qy   1345  MVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTE  1404
                |||           |            ::|      |     : |  ::|| |:
Db    703  AEATGKD------TSSITSCGDGN--VVKQELSPKKKENNAL------------LRYLLDR  744

Qy   1405  DELPGFIHGLTHGTLTPAELLQK---------GIEHQMKQGSKLTFSVDTSANLDLRAGINL  1457
                 :|:|         |:|:|||          ||||   : |     :| || ||
Db    745  DD---------PSDALSKELQPQVEGVDNKMSQCTSSTIP-SSSQEKDPKIKTET  789

Qy   1458  NEDGSKP----------NGVTARVSAGLSASANLA-AGSRE--RSTTSGQFGSTTSASNNR  1505
                 :||:            |||          :|||::     :|| : || :::||
Db    790  SEEGSGDLDNLDAILGDLTSSDFYNNSISSNGSHLGTKQQVFQGTNSLGLKSSQSVQSIR  849

Qy   1506  PTFLNGVGAGANLTAALGVAHSSTHEGKPVGIFPAF------------TSTNVSAALA  1551
                 || |          :|  :| ||              |||| :||
Db    850  PPYNRAVSLDS---------PVSVGSSPPVKNISAFPMLPKPQMLGGNPRMMDSQENYGSSMG  903

Qy   1552  LDNR---TSQSIS----------LELKRAEPVTSNDISELTSTLGKHFKDSATTKMLAAL--  1598
                :|||     :|:|          ||| || | ::|  ||| :|| |     :||  |
Db    904  GPNRRNVTVTQTPSSGDWGLPNSKAGRMEPMNSN----SMGRPGGDYNTSLPRPALGG  956

Qy   1599  KELDDAKPAEQLHILQQHFSAKDVVGDERYEAVRNLKKLVIRQQAADSHSM  1649
                          ||                                :|||
Db    957  SIPTLPLRSNSIPGARP----------VLQQ------------QQQMLQMRPG  987

Qy   1650  ELGSASHSTTYNNLSRINN----------DGIVELLHKHFDAALPASSAKRLGEMMNNDPALKDI  1704
                :||::                 |  ::|| | ||:     :|| :|||
Db    988  EIPMGMGANPYGQAAASNQLGSWPDGMI-----------SMEQVSHGTQNRPLLRNS  1033

Qy   1705  IKQLQSTPFSSASVSMELKDGLREQTEKAILD----------GKVGREEVGVLFQDRNNLR  1755
                 ||     :|  ::|    ::|          ||  :| :||
Db   1034  LDDLVGPPSNLEG------------QSDERALLDQLHTLLSNTDATGLEEID-----------  1073

Qy   1756  VKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYGQD--QNTPRR--FT  1810
                :::            || :|           : |  :|       :::|:
Db   1074  -RALGIPELVNQGQALE-PKQDAFGQGEAAVMMDQKAGL--YGQTYPAQGPPMQQGFH  1127

Qy   1811  LEGGIAQANPQVASALTDLKKEG  1833
                :||              :||
Db   1128  LQG---QSPSFNSMMNQMNQQG  1146
```

RESULT 10
PCT-US01-03782A-96
; Sequence 96, Application PC/TUS0103782A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Ford, John E et al
; TITLE OF INVENTION: Novel Bone Marrow Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-040
; CURRENT APPLICATION NUMBER: PCT/US01/03782A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19

; PRIOR APPLICATION NUMBER: 60/250,583
; PRIOR FILING DATE: 2000-11-30
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: Custom
; SEQ ID NO 96
; LENGTH: 5373
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-03782A-96

Query Match   2.0%;   Score 188.5; DB 1;   Length 5373;
Best Local Similarity   17.7%;   Pred. No. 0.071;
Matches 412; Conservative 319; Mismatches 799; Indels 803; Gaps 113;

```
Qy    27  ALQQGSS-----------SSSPQNAAASLAA----EG---KNRGKM-PRIHQPSTAADGISAAHQ  72
                :|||              ::|                :      : :::
Db  1549  ALQKNQSDLKDLQDDIQNRATSFATVVKDIEGFMEENQTKLSPR--ELTALREKLHQAKE  1606

Qy    73  QKKSFSLRGCLGTKKFSRSAPQG-QPGTTHSKGATLRDLLARDDGETQHEAAAPDAARLT  131
                :|             : ||:          :| :|    |::  :
Db  1607  QYEALQEETRVAQKELEEAVTSALQQETEKSKAA--KELA--ENKKKIDALLDWVTSVG  1661

Qy   132  RSGGVKRRN--MDDMAGRPMVKGG-----------SGEDKVPTQ-----------QKRH----QL  168
                :|          |:|                    :|           :::
Db  1662  SSGGQLLTNLPGMEQLSGASLEKGALDTTDGYMGVNQAPEKLDKQCEMMKARHQELLSQQ  1721

Qy   169  NNF---GQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPVGSTSKA-TTAHA  224
                       :||   :|  ::  |  |      :|::          :|| |:
Db  1722  QNFILATQSAQAFLDQHGHNLTPEEQOMLQQ-----------KLGELKEQ--YSTSLAQSEAEL  1772

Qy   225  DRVEIAQEDDDSEFQQLHQ----------QRLARERENPQPPKLGVATPISARFQPKLTAVAE  278
                 | :|||:: :||:        ::|            :|    :
Db  1773  KQVQTLQ-DELQKFLQDHKFEFESWLERSEKELENMHK-----------GGSSP-----------  1813

Qy   279  SVLEGTDTTQSPLKPQSML----------KG---SGAGVTPLAVTLDKGKIQLAPDNPP  324
                 :                |      :|
Db  1814  ------ETLPSLLKRQGSFSEDVISHKGDLRFVTISGQKVLDMENSFKEGK-----------EPS  1861

Qy   325  ALNTLLKQTLGKDTQHYLAHHASSD--GSQHLLLDNKGHLFDIKSTATSYSVLHNSHPGE  382
                |          ::                 :|| : :    || |
Db  1862  EIGNLVKDLKLKDATERYTALHSKCTRLGSHLNMLLGQYHQF--QNSADSLQAWMQACEAN  1919

Qy   383  IKGKLAQAGTGSVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARP  442
                 |          :    : | ::  :| |::  :: ::
Db  1920  VEKLLSD---------------------TVASDPGVLQEQLAT-----------------TKQ  1944

Qy   443  QGESIRLHD---DKIHILHPELGVWQSADKDTHSQLSRQADGKL------YALKDNRTL  492
                :||           :  |  :    : |:|              ::||: |
Db  1945  LQEELAEHQVPVEKLQKVARDIMEIEGEPAPDHRHVQETTDSILSHFQSLSYSLAERSSL  2004

Qy   493  ----------QNLSDNKSSEKLVDKIKSYSVDQRG-QVAILTD----------TPGRHKMSI--M  534
                          ::|:          ||: |                 :|::  :
Db  2005  LQKAIAQSQSVQESLESLLQSIGEVEQNLEGKQVSSLSSSGVIQKALATNMKLQODIARQK  2064

Qy   535  PSLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQ  594
                :|: | :|   :| ||    ||||:|:||:|       : :|  |
Db  2065  SSLEATREM-VTRFMETAD------------STTAAVLQGKL--AEVSQRFEQLCLQQQ  2108

Qy   595  GDGNELKMKAMPQHALDEHFGHDHQISGFFHDDHGQLNALVKNNFRQHACPLGNDHQFH  654
                | |||:|  |||  :    ||||:| || :  ::  |::: |
Db  2109  EKESSLK-KLLPQAEMFEHLS----------------GKLQQFMENKSRM-----------LASGNQ-  2147

Qy   655  PGWNLTDALVIDNQLGLHHTNPEPH----------EILDMGHLGSLALQEGKLHY----FD  701
                 | | ||     :   || |             :| |       :|:|      :
Db  2148  PDQDITHFQIQELNLEMEDQQENLDTLEHLVTELSSCGFALDLCQHQDRVQNLRKDFT  2207

Qy   702  QLTKGWTGAESDCKQLKKGLDGAAYLLK-----------DGEVKRLNINQSTSSIKHGTENV  752
                 :| |        ::|     |:       :| :|  :|::|  :
Db  2208  ELQKTVKEREKDASSCQEQLDEFRKLVRTFQKWLKETEGSIPPTETSVSAKELEKQIEHL  2267

Qy   753  FSLPHVRNKPEPGDALQLNKDKAQAMAVIGVNKYLALTE----------KGDIRSFQIKPGTQ  807
                | ||  | |      ::      ::    |  |:                 |:
```

```
Qy  1499  TSASNNRPTFLNGVGAGANLTAALGVAHSSTHE-GKPVGIFPAFTSTNVSAALALDNRTS 1557
                :  ::  :  :::  :::  ::     :    ::  ::
Db  3305  ESLESRWTELLSKAAARQKQLEDILVLAKQFHETAEPISDFLSVTEKKLANSEPVGTQTA 3364

Qy  1558  --QSISLELKRAEPVTSN------------------DISELTSTLGKHFKDSATTKMLAALKELD 1602
              :  :   :::  :                   :::  :::  :: ::::
Db  3365  KIQQQIIRHKALEEDIENHATDVHQAVKIGQSLSLSLTSPAEQGVLSEKIDSLQARYSEIQ 3424

Qy  1603  DAKPAEQLHILQQHFSAKDVVGDERYEA------------------VRNLKKLVIRQQAAD-- 1645
            :  ::  :   :                        :  :::  : : :
Db  3425  D-RCCRKAALLDQALSNARLFGEDEVEVLNWLAEVEDKLSSVFVDFKQDVLHRQHADHL 3483

Qy  1646  ------------------SHSMELGSA-SHSTTYNNLSRINN--DGIVELLHKHFDAALPASSAK 1689
                            ::      :    ::: :::     :::  :: ::
Db  3484  ALNEEIVNRKKNVDQAIKNGQALLKQTTGEEVLLIQEKLDGIKT--RYADITVTSSKAL 3540

Qy  1690  RLGEMMNNDPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAIILDGKVGREEVG--V 1746
           :                              :::  : :::  ::
Db  3541  RTLEQ----------------YEELTGWLREVEEELATSG--GQSPTGEQIP 3585

Qy  1747  LFQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKY 1799
                :  :::   :  :::  :::  ::  :::       :::
Db  3586  QFQQRQKELKKEVMEHRLVLDTVNEVSRALLELVPWRAREGLDKLVSDANEQY 3638
```

RESULT 11
US-08-945-567C-4
; Sequence 4, Application US/08945567C
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Ken
; APPLICANT: HARKNESS, Robin E.
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: CHONG, Pele
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE PROTEIN OF
; TITLE OF INVENTION: MORAXELLA
; FILE REFERENCE: 1038-745 MIS
; CURRENT APPLICATION NUMBER: US/08/945,567C
; CURRENT FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 08/431,718
; PRIOR FILING DATE: 1995-05-01
; PRIOR APPLICATION NUMBER: 08/478,370
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/621,944
; PRIOR FILING DATE: 1996-03-26
; PRIOR APPLICATION NUMBER: PCT/CA96/00264
; PRIOR FILING DATE: 1996-04-29
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1833
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-08-945-567C-4

```
Query Match             2.0%;   Score 184.5;  DB 4;   Length 1833;
Best Local Similarity  18.3%;   Pred. No. 0.018;
Matches   284;  Conservative  214;  Mismatches  537;  Indels  521;  Gaps  71;

Qy   539  ASPESHISLSLHFADAHQG--LLHGKSELEAQSVAISHGRLVVADSEGK--LFSAAIPK 593
            ::     :::  : :       :  ::  ::   ::  :::
Db    16  ATAKSAYSLAVGLAATAEGQSTIAIGSDATSSSLGAIALGAGTRAQLQGSVIALGQGSVVT 75

Qy   594  QGDGNELKMKAMPQHALDEHFGHDHOISGFFHDHGQLNALVKNNFRQQHACPLGNDHQF 653
             :::  :: :  :::  : :::       :::     :
Db    76  QSDNNSRPAYTPNTQALDPKFQATNNTKA-----------GPLS--IGSNSIKRKIINVG-- 122

Qy   654  HPGWNLTDALVI------------------DNQLGLHHTNPEPHEILDMGHLGS 689
           : :  :: :::        :  :::  :  :
Db   123  -AGVNKTDAVNVAQLEAVVKWAKERRITFQGDDNSTDVKIGLDNT---------------- 166
```

```
Db  2268  KSL--LDDWASKGTLVEEIN--CKGTSLE----NLIMEITAPDSQGKTDLTEIQCDMSDV 2319
            :::      :  :  :::        :::  :::
Qy   808  QL--ERPAQTL-----SREGISGELKDIHVDHKQNLYALTHEGEV--------------- 845
Db  2320  NLKYEKLGGVLHERQESLQAILNRMEEVHKEANSVLQWLESKEEVLKSMDAMSSPTKTET 2379
            :::       :  :::  ::       :::
Qy   846  --------------FHQPREA---------WONGAESSSWHKL-----------AL 867
Db  2380  VKAQAESNKAFLAELEQNSPKIQKVKEALAGLLVTYPNSQEAENWKKIQEELNSRWERAT 2439
             :  :::  ::: :::      :    :::      : :::  :::
Qy   868  PQSESKIKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPER----GPLAV------- 915
Db  2440  EVTVARQRQLEESASH--LACFQ-AAESQLRP--W----LMEKELMMGVLGPLSIDPNML 2490
            :::  :::   :::      :        : :::
Qy   916  GTSGSQTVF----------------NRLMQGVKGKVIPGSG------------------L 941
Db  2491  NAQKQQVQFMLKEFEARRQQHEQLNEAAQG-ILTGPGDVSLSTSQVQKELQSINQKWVEL 2549
            ::: :::  :::       :  :::     :::  :::
Qy   942  TVKLSAQTGGMTGA--------EGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGWQG 996
Db  2550  TDKLNSRSSQIDQAIVKSTQYQELLQDLSEKVRAVG-------QRLSVQSAISTQ----- 2597
            : :::    :        ::: :::       :::
Qy   997  REGLKPLYEMQGAL---IKQLDAHNVRHNAP------------------QPDLQSKLETLDIGE 1039
Db  2598  PEAVKQQLEETSEIRSDLEQLD-HEVKEAQTLCDELSVLLIGEQYLKDELKKRLETVALPL 2656
              ::  :::       :    :::  :::     ::  : ::: :::
Qy  1040  HGAE---------------------------LLNDMKRFRDELEQSATRSVTV--------- 1065
Db  2657  QGLEDLAADRINRLQAALASTQQFQQMFDELRTWLDDKQSQQAAKNCPISAKLERLQSQLQ 2716
             :::       ::  :::      :::  : :::
Qy  1066  ----------LGQHQG---VLKSNGE-INSEFKPSPGKALVQS--FNVNRSGQDLSKS---- 1107
Db  2717  ENEEFQKSLNQHSGSSYEVIVAEGESLLLSVPPGEEKRTLQNQLVELKNHWEELSKKTADR 2776
            :::  : :::      :  :::       :  :::
Qy  1108  ------------LQQA-----------------VHATPPSAESKL--QSMLG---- 1128
Db  2777  QSRLKDCMQKAQKYQWHVEDLVPWIEDCKAKMSELRVTLDPVQLESSLLRSKAMLNEVEK 2836
            : :::       :::  :  ::  :::    ::  :::
Qy  1129  ------HFVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLILDTVT-----IGELHELA 1177
Db  2837  RRSLLEILNSAADILINSSEADEDGIRD--EKAGINQN--MDAVTEELQAKTGSLEEMT 2891
            :::       :  :::  : :::      :::
Qy  1178  DKAK------------------------------------------------ 1181
Db  2892  QRLREFQESFKNIEKKVEGAKHQLEIFDALGSQACSNKNLEKLRAQQEVLQALEPQVDYL 2951
              ::  ::: :::     :  :::
Qy  1182  --LVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMG--FTHNK--------- 1225
Db  2952  RNFTQGLVED-APDGSDASQLLHQAEVAQQEFLE--VKQRVNSGCVMMENKLEGIGQFH 3007
             :::  :::     ::  :::      : :::  :::
Qy  1226  -----------ALEANYDAVKAFIN---AFKKEHHGVNLTTRT 1254
Db  3008  CRVREMFSQLADLDDELDGMGAIGRDTDSLQSQIEDVRLFLNKIHVLKLDIEASEAECRH 3067
            ::  :::       :  :::  : :::     :::
Qy  1255  VLESQGSAEL-------------------AKKLKNTLLSLDSGESMSFSRSYGG--- 1289
Db  3068  MLEEEGTLDLLGLGKRELEALNKQCGKLTERGKARQEOLELTLGRVEDFYRKLKGLNDATT 3127
            ::  ::: :::     :  :::  :::      :    :::
Qy  1290  -------GVSTVFV--------PTLSKKVPVPVIPGAGITLDRAYNLSFSRTSG--- 1328
Db  3128  AAEEAEAALQWVVGTEVEIINQQLADFKMFQKEQVDPL------------QMKLQQVNGLG 3175
            ::: :::     :  :::      :::  :::
Qy  1329  -GLNVSFGRDGGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKIS---PDLRIGA 1384
Db  3176  ----LEHDMEEI----------NAR-WNTLNKKVAQRIAQLQEAL 3220
            : :::      :::  :::
Qy  1385  AVSGTLQGTLQNSLKFKLTEDEL----PGFIHGLTHGTLTPAELLQKGIEHQMKQGSK 1438
Db  3221  LHCGKFQDALEPLLSWLADTEELIANQKPPSAEYKVVKAQIQEOKLLQRLLD------- 3272
            :::      :::  : :::       :  :::
Qy  1439  LTFSVDTSANLDLRAGINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGST 1498
Db  3273  -DRKATVDM-----LQAEGGR----------IAQSAELA-DREKIT------GQL 3304
```

us-09-596-784-2.rapn

```
Qy    690  LALQEGKLHYFDQLTKGWTGA--ESDCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIKH  747
                :|:||  |  |||  :||  :: |||:||  ||:|:|:  |:
Db    167  LTIKGGA--ETNALTDNNIGVVKEADNSGLKVKL--AKTLNNLTEVNTTTLN-ATTTVKV  221

Qy    748  G-----TENVFSLPHVRNKPEPGD--------ALQGLNKDDKAQAMAVIGVNKYLALTEK  794
           |     ||  |      |        ||:  : :  :|
Db    222  GSSSSTTAELLSDSLIFTQPNTGSQSTSKTVYGVNGVKFTNNAETTAAIGTTR-------  274

Qy    795  GDIRSFQIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFHQPREAWQ  854
              |||:   |      | : |:  |:
Db    275  --------ITRDKI-GFARDGDVDEKQAPY-------------------------------  295

Qy    855  NGAESSSWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLA  914
                         |:  ||  |:          |:||       |
Db    296  --------LDKKQLKVGSVAI-------TIDNG-----IDAGN------KKISNLA      325

Qy    915  VGTSGSQTVFNRLMQGVKGKVIPGSGLTV-------KLSAQTGGMT----------GAE  956
           ||:|                |  ||  ||        | ||:                ::
Db    326  KGSSANDAVTIEQLKAAKPTLNAGAGISVTPTEISVDAKSGNVTAPTYNIGVKTTELNSD  385

Qy    957  GRKVSSKFSERIRAYAFNPTMSTPRPI                                  983
           | :   |||       | |
Db    386  G--TSDKFS--VKGSGTNNSLVTAEHLASYLNEVNRTADSALQSFTVKEEDDDANAITV   441

Qy    984  ----KNAAYATQHGWQGREGLKPLYEMQGALIKQL------------DAHNVRHNA     1023
               ||  || |    |:  |             ||            |
Db    442  AKDTTKNAGAVSILKLKGKNGLTVATKKDGTVTFGLSQDSGLTIGKSTLNNDGLTVKDTN  501

Qy   1024  PQPDLQS--KLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHOGVLKSNGEIN  1080
               |    |||| ||||  |||||   || ||:|:|    :    |      || ::|
Db    502  EQIQVGANGIKFTNVNGSNPGTGIANTARITRDKIGFAGS--------DGAVDTNKPYL   552

Qy   1081  SEFKPSPGKALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQ  1140
                 ||   | |    :|      |:|| |||                |:|
Db    553  DQDKLQVGNVKITNTGINAGGKAIT-----GLSPTLPSI----------ADQSSR      592

Qy   1141  KGEIPLGRQRDPNDKTALTKSRLILDT-------------VTIGEL              1173
            ||||||         :::  |   ||                |
Db    593  --NIELGNTIQDKDKSNAASINDILNTGFNLKNNNNPIDFVSTYDIVDFANGNATTATVT  650

Qy   1174  HELADK-AKLVSD--------HKPDADQIKQLRQOFDTLREKRYESNPVKHYTDMGFTHN  1224
            |  ||  :|:||        || | |  |    |:  |   |:         | |
Db    651  HDTANKTSKVVYDVNVDDTTIHLTGTDDNKKLGVKTTKLNKTSANGN-----TATNFNVN  705

Qy   1225  KALE---ANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLLSLDSGESM  1281
            |||    :| | |  |:   :|     |     |  :    |:|::|
Db    706  SSDEDALVNAKDIAENLNTLAKEIHTTKGTADTALQT----FTVKKVDENNNADDANAIT  761

Qy   1282  SFSRSYGGGVSTVFVFPTLSKKVPVPVIPGAGITLDRAYNLSFS-RTSGGLNV--SFGRDG  1338
            | ||:     ||| |    |:    | |||| |||: ||| ::  ||:||
Db    762  VGQKNANNQVNTL--TLKGE-------NGLNIKTDKNGTVTFGINTTSGLKAGKSTLNDG   812

Qy   1339  GVS------GNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRIGAAVSGTLQGT  1393
            |:|       |::   | |    :  |||:   |::::|   :|| | ||||:|||: ||
Db    813  GLSIKNPTGSEQIQVGADGVKFA----------KVNNNGVVGAGIDGTT---          851

Qy   1394  LQNSLKFKLTEDELPGFIHGLTHGTLTPAE--LLQKGIEHQMKQGSKLTFSVDTSANLDL  1451
            :         ::|   || ||  ||   |   ::|:||  |  :|   | ::|  ::
Db    852  ------RITRDEI-GFTG--TNGSLDKSKPHLSKDGIN--AGGKKIT------NI       889

Qy   1452  RAG-INLNE---DGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTSASNNRP  1506
            ||  :||    | |||||||| | ||:|||||||||   ||||||||||||||||| |
Db    890  QSGEIAQNSHDAVTGGKIYDLKTELENKISSTAKTAQNSLHEFSVADEQGNNFTVSNPYS  949

Qy   1507  TF------LNGVGAGANJTAALGVAHSSTHEGKPVGIFPAFTSTNVSAALALD        1553
            ||      | ||||||| |||||||||||||||||||:|||||||||||:|:|:
Db    950  SYDTSKTSDVITFAGENGITTKVN---KGVVRVGIDQTKGLTTPKLTVGNNNGKGIVID  1005

Qy   1554  NRTSQSISLELKRAEPVTSNDISELTSTLGKHFKDSA--TTKMLAALKELDDAKPAEQL  1610
                  |:|||| | |||:|||:|||||||||||:|   ||||||||||:||||||||||
Db   1006  SQNGQ-------NTITGLSNTLANVTNDKGSVRTTEQGNIIKDEDKTRAASIV        1051

Qy   1611  HILQQHFS----AKDVVGDERYEAVRNLKKLVIRQQAADSHSMELGSASHSTTYNNLS  1664
```

```
Db   1052  DVLSAGFNLQGNGEAVDFV--STYDTV---------------------ATTAKVTYDDTS  1094
           :||  |  :|||  || |   :|:||                       |||||:| |||
Qy   1665  RIN------NDGIVELLHKHF------------------------DAALPAS----    1686
           ||        |||:|||| |                        |||||
Db   1095  KTSKVVYDVNVDDTTIEVKDKKLGVKTTLTSTGTGANKFALSNQATGDALVKASDIVAH  1154

Qy   1687  --------SAKRLGEMMNN----DPALKDIIKQLQSTPFSSASVSMELKDGLREQTEKA  1733
                   |||:| :||     || |||||||||||| :|::|:| |||||||:| ||:|
Db   1155  LNTLSGDIQTAKGASQANNSAGYVDADGNKVIYDSTDNKYYQAK--NDGTVDKTKEV    1209

Qy   1734  ILDGKVGREEV--GVLFQDRNNLRVKSVSVSQSV--SKSEGFNTP-ALLGTSNSAAMS  1787
           |||||||:||   |:||||||||:|||||||||    ||||||| ||||||||||||||
Db   1210  AKDKLVAQAQTPDGTLAQ-----MNVKSVINKEQVNDANKKQGINEDNAFVKGLEKAASDN  1265

Qy   1788  MERN-----IGTIN----------FKYGQDQNPRR----FTLEGGIAQANPQVASALTD  1828
           |:|      ||:|            |||||||| ||    |||||||||||||||||||||
Db   1266  KTKNAAVTVGDLNAVAQTPLTFAGDTGTTAKKLGETLTIKGGQTDTN----KLTD    1316
```

```
Query Match            1.8%;    Score 167;    DB 1;    Length 2189;
Best Local Similarity 18.8%;    Pred. No. 0.32;
Matches  373;    Conservative  224;    Mismatches  754;    Indels  634;    Gaps  91;

Qy     89  SRSAPQGQPGT------THSKGATLRDLLARDDGETQHEAAAPDAARLTRSGGVKRRNMD   142
           |:: |      ||
Db    412  SAEAPKGSPGSWWKKELSGSSSAPKLEYTVRTDTQSPTNTGSPSSPQ-QKSEGLGSRHRP   470

Qy    143  DMAGRPMVKGGSGEDKVPTQQ-----------KRHQLNNF--------GQMRQT        177
                 | :| | |   :             |||  ||
Db    471  VARVSPHCKRSEAEAKPSGSQTVNLTGRANDPCDLLDSRVQATSVKVTVAGFQPGGAVEKE   530

Qy    178  MLSKM------------AHPASANAGDRLQHSPPHIPGSHHEIKEEP-------VGSTS    217
            ||:|            ||                              |||:
Db    531  SLGKLJTTGDACVSTCELASALSHLDASHLTENLPKAASELGQOPMTELDSSSDLISSPG   590

Qy    218  KATTAHADRVEIAQEDDDSEFQQLHQQRLARRERENPPQP--PKLG----VATPISAR    268
```

```
Db   591 KKGAAHPDPSKTSVDTG------KVSRP-ENPSQPASPRVAKCKARSPVRLPHEGS       639
           ||| |              ::  |||  |||  |     ::|
Qy   269 FQPKLTAVAE--SVLEGTDTTQSP-LKPQSMLKGSG--AGVTPLAVTLDKGKLQLAP----P 320

Db   640 PSPGEKAAAPPDYSKTRSASETSTPHNTRRVAALRGAGPGAEGMTPAGAVL------P     691
            |        |||| |               | | ::   |  :  |      ::| |
Qy   321 DNPPALNTLLKQTLGKDTQHYLA--HHASSDGSQHLLLDNKGHLFDIKSTATSYSVL       375

Db   692 GDP--LTSQEQRQAPGPGNHSKALEMTGIHAPESSQEPSLLEGADSV-SSRAPQASLSML   747
          | |                    |  |  |  |||            |             ||
Qy   376 HNSH------PGE------IKGKLAQAGTGSVSVDGKSGK----                    403

Db   748 PSTDNTKEACGHVSGHCCPGGSRESPVTDIDSFIKELDASAARSPSSQTGDSGSQEGSAQ   807
          |                   ||       ||  |  ||          ||
Qy   404 ----ISLGSGTQSHNKTML-SQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHILH   458

Db   808 GHPPAGAGGGSSCRAEPVPGGQTSSPRRAWAAGAPAYPQWASQP------              851
             |                  ||| :             :     :|
Qy   459 PELGVWQSADKDTH------SQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSV  513

Db   852 ---SVLDSINPDKHFTVKNFLSNYSRNFSSFHEDSTSLSGLGD--STEPSLSSMYGDAE     906
             ::||                 :  | ||       :|   :  ||   |    :  |
Qy   514 DQRGQVAILTDTP---GRHKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQSV   570

Db   907 DSSSDPESLTEAPRASARDGWS------PPRSRV--SLHKED------PSESEEEQI      949
           | :   |   |                   |||                ||
Qy   571 AISHGR------LVVADSEGKLF------SAAIPKQ------GDGNEL               600

Db   950 EICSTRGCPNPPSSPAHLPTQAAICPASAKVLSLKYSTPRESVASPREKAACLPGSYTSG 1009
          :                      |       |                          NALV 635
Qy   601 KMKAMPQHALDEHFGHDHQISGFFHDDHGQL------

Db  1010 PDSSQPSSLL------EMSSQEHETHADISTSQNHRPSCAEETEVTSASSAMENSPL     1061
            | :                 ||               |
Qy   636 KNNFRQQHACPL-----GNDH------QFHPGWNLTDAL--------V               664

Db  1062 SKVARHFHSPPIILLSSPNMVNGLEHDLLDDETLNQYETSINAAASLSSFSVDVPKNGESV 1121
          :                       | |          :
Qy   665 IDNQLGLHHTNPEPHEILDMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGA  724

Db  1122 LEN----LHIS--ESQDLDDLLQKPKMIARRPIMAWFKEINKHNQGTHLRSKTEKE---- 1171
          |        ::    :                    |  :
Qy   725 AYLLKDGEVKRLNINQSTSSIKHGTENVFSLPHVRNKPEPGDALQG--LNKDDKAQAMAV  782

Db  1172 -QPLMPARSPDSKIQMVSSSQKKGV----TVPH--SPPQKTNLENKDLSKKSPAEMLLT  1224
             |                            |||
Qy   783 IGVN------KYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDI--HVDH    832

Db  1225 NGQKAKCGPKLKRLSLKGKAKVNS------EAPAANAVAKAGGTDHRKPLISPQTSH    1274
           |                              ||
Qy   833 K------QNLYALTHEGEVFHQPREAWQNGAESSSWHKLALPQS------ESKLKSLDMS  880

Qy  1275 KTLSKAVSQRLHVADH------EDPDRNTTAAPRSPQCVLESK-PPLATS           1317
            :                     |    |               |
Qy   881 HEHKP------IATFED--GSOHQLKAGGWHAY----AAPERGPLAVGTSGSQTVFNR    926

Db  1318 GPLKPSVSDTSIRTFVSPLTSPKPVPEQGMWSRFHMAVLSEPDRGCPTTPKSPKCRAEGR 1377
            |                    ::                   :    |   |
Qy   927 LMQGVKGKVIPGSGLTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPI--K  984

Db  1378 APRADSGPVSP------AASRNGMSVAGNRQSEPRLASHVAA----DTAQPRPTGEK    1424
           |     |                   ||                    | |
Qy   985 NAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAP------QPDLQSKLE       1033

Db  1425 GGNIMASDRLERTNQLK-IVEISAEAVSE---TVCGNKPAESDRRGGCLAQGNCQEKSE  1479
          |  :                 :                   |
Qy  1034 T---LDLGEHGAELLNDMKRFRDELEQSATR--SVTVLGCHQGVLKSNGEINSEFKPSPG 1088

Db  1480 IRLYRQVAESSTSHPSSLPSHASQAEQEMSRSFSMTKLASSSSSIQT-AIRKAEYSQGKS 1538
          :  ::                |              |          |
Qy  1089 KALVQSFNVNRSGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSH--QKGEIP- 1145
```

```
Db  1539 SLMSDSRGVPRNSIPGGPSGEDHLYFTPRPA-TRTYSMPAQF------SSHFGREGHPPH 1591
          :    :                                          :|
Qy  1146 -LGRQRDPNDKTALTKSRLILDTVTIGELHELA------DKAKLV                 1183

Db  1592 SLGRSRD--SQVPVTSSVVPEAKASRGGLPSLANGQGIYSVKPLLDTSRNLPATDEGDII  1649
          |  |                | |    | | |  || ::
Qy  1184 SDHKPD--ADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINA   1240

Db  1650 SVQETSCLVTDKIKVTRRHY------CYEQN------WPHE--STSFFSVKQRIKS      1691
          :  |             |            ||          |     | ::  |
Qy  1241 FKKEHHGVNLTT--RTVLESQGSAELAKKLKNTLLSLDSGESMSFSRSYGGVST-       1293

Db  1692 FE------NLANADRPVAKSGASPFLSVSSKPPIGRRSGSIVSGSLGHPGDAAARLLRR   1745
          |:                   |  || :           |
Qy  1294 ---VFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGRDGGV            1340

Db  1746 SLSSCSENQSEAGTLLPQMAKS--PSIMTLITISRONPPETSSKGSDSELKKSLGPLGIP  1802
          :              |                    |:               |
Qy  1341 SGNIMVATGHDVMPYMTGKKT------SAGNAS                             1367

Db  1803 TPTMTLAS---PVKRNKSSVRHTQPSPVSRSKLQELRALSMPDLDKLCSEDYSAGPSA    1857
           | |         ::             |       |              | ||
Qy  1368 DWLSAKHKISPDLRIGAAVSGTLQGTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQK 1427

Db  1858 VLFKTELEITPRRSPGPPAGGV------SCPEKGGNRACPGGSGPKTSAAETPSSASDT   1910
          ::           :|   || :             |
Qy  1428 GIEHQMKQGSKLTFSVDTSANLD------LRAGIN                           1468

Db  1911 G------EAAQDLPFRRSWSVNLDQLLVSAGDQQRLQSVLSVGSKSTILTLIQEAKAQSE 1965
          |           :|  :  |       |       | |
Qy  1469 ------ARVSAGLSASANLAAGSRERSTTSGQFGSTTSASN------NRPTFLNGVG    1513

Db  1966 NEEDVCFIVLNRKEGSGLGFSVAGGTDVEPKSITVHRVFSQGAASQEGTMNRGDFLLSVN 2025
          :      | |       :| |             |   | | |         |
Qy  1514 AGANLTAALGVAHSST------HEGKPVGIFPAFTSTNVSAALALDNRTSQSISLEL     1564

Db  2026 -GASLA--GLAHGNVLKVLHQAQLHKDALVVIKKGMDQPRPSA                   2065
             || |:     ::     :  |          |
Qy  1565 KRAEPVTSN------DISELTSTLGKHF--KDSATTK                         1593

Db  2066 -RQEPPTANGKGLLSRKTIPLEPGIGRSVAVHDALCVEVLKTSAGLGLSLDGGKSSVTGD  2124
           :|||   |           :          |      | |
Qy  1594 MLAALKELDDAKPAEQLHILQQ------HFSAKDVGDERYEAVRNLK------KLVIRQ   1641

Db  2125 GPLVIKRVYKGGVAEQAGIIEAGDEILAINGKPLVGLMHFDAWNIMKSVPEGPVQLLIRK  2184
          ||
Qy  1642 QAADS                                                         1646

Db  2185 HRNSS                                                         2189
```

RESULT 13
US-09-739-449-8806
; Sequence 8806, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 8806
; LENGTH: 1228
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-09-739-449-8806

Query Match   1.7%;   Score 164;   DB 5;   Length 1228;
Best Local Similarity   18.6%;   Pred. No. 0.17;

```
Matches 282; Conservative 185; Mismatches 513; Indels 536; Gaps 65;

Qy   488 DNRTLQNLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTPGRHKMSIMPSLDASPESHISL  547
Db    28 DRSTLWNAAE-----FAEKRKDARVAREFEIAL------PHELSPEGRLKAARAFAQDL   75

Qy   548 SLHFADAHQGLLHGKSELEAQSVAISHGRLVVADSEGKLFSAAIPKQGDGNELKMKAMPQ  607
Db    76 ANRYGAAVDFAIHSPSE--------HGDIRNHHAHVLMTTRQVGKAGLGEKTCL-----  121

Qy   608 HALDEHFGHDHQISGFFHDHGQLNAL------VKNNFRQQHACPLGNDHQFHPGWNLTD  661
Db   122 --EHKNARLLANGMATTDM-QLRDIRQSWEGIANRQLQHEGLDVRIDHRSH-------  169

Qy   662 ALVIDNQLGLHHTNEPHEILDMGHLG--------SLALQEGKLHYFDQLTKGWTGAES  712
Db   170 -IERGLELSPTE------HMGVHASQMRQQGMAVERGRL----------  201

Qy   713 DCKQLKKGLDGAAYLLKDGEVKRLNINQSTSSIKHGTENVFSL-PHVRNKPEPGDALQGL  771
Db   202 ------DDEAAR----QNAALIRQKPEQVLTLISHEKSVFDRHDIAKTL  240

Qy   772 NK--DDKAQ---AMAVIGVNKYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGEL  825
Db   241 HRYINDDAQTFQNAFAAV-----LASSALVELQAERIDPGTGKVSN-ARYSTREMIDLEL  294

Qy   826 KDIHVDHKQNLYALTHEGEVFHQPREAWQNGAESSWHKLALPQSES-KLKSLDMSHEHK  884
Db   295 ----AMARSAVRLHQ--AQSHGVDPRHVDRAIERQDRSLRRSSGGMLAASD  339

Qy   885 PIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGS-QTVFN--RLMQGVKGKVIPGSG  940
Db   340 PSAGLSDEQRHAIK-----HITGSERIAVVVGFAGAGKSTMLTAARKAWEAQGYQVHGAA  394

Qy   941 LTVKLSAQTGGMTGAEGRKVSSKFSERIRAYAFNPTMSTPRPIKNAAYATQHGW-QGREG  999
Db   395 LSGK------AAEGLEESSGIESRTLA----------SWSYSWDQGR--  425

Qy  1000 LKPLYEMQGALIKQLDAHNVRHNAPQPDLQSKLETLDLGEHGAELLNDMKRFRDELEQSA 1059
Db   426 --------------------NLIGSSDLFVIDEAGMVGSRQLARFIGEAEERG  458

Qy  1060 TRSVTVLGQHQGVLKSNGEINSEFKPSPGKAL---VQSFNVNRSGQDLSKSLQQAVH 1113
Db   459 AK-IVLVGDHE------QLQAIGAGAPFRAIAEQIGHVELSGIRRQRHDWQRQASVA-F  509

Qy  1114 ATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTALTKSRLILDTVTIGEL 1173
Db   510 ATHKTAE---------GLAAYRDHGDIHFAESRD---------  534

Qy  1174 HELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDMGFTHNKALEANYDA 1233
Db   535 ---AAMAQIVRDYIDD-----------NEKRPDG----TRVAMAHRR---ADVRA  568

Qy  1234 VKAFINAFKKEHH-----------GVNLTTRTVLESQGS---AELAKKLKNTLLSLDSGE 1279
Db   569 LNATIRSELQNRQRLERSLGLSDGDPRGDRGDVEDRGNSGDVAELTFQTSNGKRAFASGD  628

Qy  1280 SMSF---------------------SRSYGGGVSTVFVPTLSKKV----PVPVI 1308
Db   629 RIIFLENNRDLGVKNGMLGTVEDVEKGRIVARLDGRGGDSVSIPTDSYQAIDHGYAVTIH  688

Qy  1309 PGAGITLDRAYNLSFSRTSCGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKK-TSAGNAS 1367
Db   689 KNQGATVDRAFVLASST-------MDRHLAYYAMTRHRDSVQLYADIKEFTSAGRLV  738

Qy  1368 DWLSAKHKISP----------DLRIGAAVSGTLQGTLQNSLKFKLTEDELP-GFIH 1412
Db   739 D----HGVAPYEHNRRARENYFVTLENDKGERHTVWGVDLN---RAMRDASPAIGDKI  789

Qy  1413 GLTHGTLTPAEL---------LQKGIE-HQMKQGSKLTFSVDTSANLDL------RA 1453
Db   790 GLQHVGATPVTLPDGTKAERNAWRVVKGEELAYQKLASRLSRSGAKETTLDYISEFAERR  849

Qy  1454 GINLNEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTSASNNRPT------ 1507
Db   850 GIASDPRIAGDQGIADQL--GLKSEIELASARKDRQEVLSRAAHLQREQQDRPSERQQVY  907

Qy  1508 ------------------------------FLNGVGAGANLTAA 1521
Db   908 EERAGDLAGPVRRGHPRSPFARDIDEGQQDRAEETNRYRIRWSDLMSQDGAVPPPTET  967

Qy  1522 LGVAHSSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI-SLELKRAEPVTSNDISELTS 1580
Db   968 VG-QQSSLETGRPTPLVPAIT--------RYQRSIEEVAQQRALSVIDQQFDTVES 1014

Qy  1581 TLGKHFKDSA-TTKMLAALKELDDAKPAEQLHILQQHFSAKDVVGD-ERYEAVR---- 1632
Db  1015 LVRRVFRDPAEVVAARLRAAMTEKEG---------NGKIMAKAMAGQPERFGELRGESGL 1064

Qy  1633 --NLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFDAALPASSAKR 1690
Db  1065 FGSNKERKEALQYARSLSAHIGYVSEAW----------ERR 1095

Qy  1691 LGEMMNNDP---ALKDIIKQLQSTPFSSASVSMELKDGLREQTEKAILDGKVGREEVGVL 1747
Db  1096 LGEERQSEQWQREQRDVIEVPGLTP-RSAEILAKVEEMQVEKRGKFIKD----- 1143

Qy  1748 FQDRNNLRVKSVSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYCQDQNTPR 1807
Db  1144 ----LR----SSAEG----QAALDEARQVADALTRRFGS-------SDPR 1174

Qy  1808 RFTLEGGIAQANPQVA 1823
Db  1175 RFAEE---LEARPELA 1187
```

RESULT 14
PCT-US01-04098A-3647
; Sequence 3647, Application PC/TUS0104098A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-029
; CURRENT APPLICATION NUMBER: PCT/US01/04098A
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: Not Yet Assigned
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO 3647
;   LENGTH: 1654
;   TYPE: PRT
;   ORGANISM: Homo sapiens
PCT-US01-04098A-3647

Query Match   1.7%; Score 163.5; DB 1; Length 1654
Best Local Similarity 18.7%; Pred. No. 0.32;
Matches 322; Conservative 198; Mismatches 606; Indels 599; Gaps 82;

```
Qy   17  AAH-NPVGHGVALQQGSSSSPQNAAASLAAEGKNRG---KMPRIHQPS-----------  61
             |||  ||    :|    :|    ::     |
Db   59  AAHPDSKTSVDTGQVSRPENPSQPASPRVTKCKARSPVRLPHEGSPSGPEKAAAPPDYS  118

Qy   62  --TAADGISAAHQQKKSFSLRGC------------LGTKFKSRSAPQGQPGTTHSKGA  105
           :|||       |    ||||
Db  119  KTRSASETSTPHNTRRVAALRGAGPGAEGMTPAGAVLPGDPLTSQEQRQGAPG-NHSKAL  177

Qy  106  TLRDLLARD------DGETQHEAAAPDAA-RLTRSGGVKRRNMDDMAGRPMVKGGSGE  156
           |         ::|       ::
Db  178  EMTGIHAPESSQEPSLLEGADSVSSRAPQASLSMLPSTDNTKEACGHVSGH-CCPGGSRE  236

Qy  157  DKVPTQQKRHQLNNFGQMRQTMLSKMAHPASANAGD------RlQHSPPHIPGSHHEIKE  210
              :|      |
Db  237  SPV----TDIDSFIKELDASAARSPSSQTGDSGSQEGSAQGHPPAGAGGGSSCRA  287

Qy  211  EPV--GSTSKATTA----------------------------------HA  224
         |||  ||
Db  288  EPVPGGGQTSSPRRAWAAGAPAYPQWASQPSVLDSINPDKHFTVNKNFLSNYSRNFSSFHE  347

Qy  225  DRVEIA----------QEDDDSEFQQLHQ--------------------QRLARERENP-  253
         |        |
Db  348  DSTSLSGLGDSTEPSLSSMYGDAEDSSDPESLTEAPRASARDGWSPPRSRVSLHKEDPS  407

Qy  254  ----------PQPPKLGVATPISARFQP----------------------KLTAV  276
                                                            |
Db  408  ESEEEQIEICSTRGCPNPPSSPAAHLPTQAAICPASAKVLSLKYSTPRESVASPREVACL  467

Qy  277  AESVLEGTDTTQSPLKPQSMLKGSG-------AGVTP----------LAVTLDKGKLQLAPDN  322
          |:         ::|
Db  468  PGSYTSGPDSSQ----PSSLLEMSSQEHETHADISTSQNHRPSCAEETTEVTSASSAMEN  523

Qy  323  PPALNTLLKQTLGKDTQHY-----LAHHASSDGSGHLLDNKGHLLFDIKSTATSYSVLH  376
           :|      ::      :|
Db  524  SP------LSKVARHFHSPPIILSSPNMVNGLEHDLLDDE------TLNQYETSI  566

Qy  377  NSHPGEIKGKLAQAGTGSVSVDGKSISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHP  436
         |
Db  567  N------AAASLSSFSVD---------VPKNGESVLENLHISESQDLDDLLQKP  605

Qy  437  AGAAR----------PQGESIRLHDDKIHILHPELGVWQSADKDTHSQLSRQADGKL  483
Db  606  KMIARRPIMAWFKEINKHNQGTHLRSKTEKEQPLMPA--RSPDSKIQMVSSSQKKGVT  661

Qy  484  Y----ALKDNRTLQNLSDNKSSEKLV-----------DKIKYSVSVDQRGQVAILTDTPGR-  528
Db  662  VPHSPPQPKTNLENKDLSKKSPAEMLLTNGQAKCGPLKRLSL--KGKAKVNSEAPAAN  719

Qy  529  ----HKMSIMPSLDASPE-SHISLS------------LHFAD------------AHQGLLH  560
Db  720  SPQTSHKTLSKAVSQRLHVADHEDPDRNTTAAPRSPQCVLE  774

Qy  561  GKSELEAQSVAISHGRLVVADSEGKLFSA------------AIPKQGDGNELKMKAMPQHALDEHF  614
          :      :|:  :
Db  775  SKPPLATSGPL----KPSVSDTSIRTFVSPLTSPKPVPEQGMWSRFHMAVLSE------  823

Qy  615  GHDHQISGFFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDN----QL  669
Db  824  ----------PDRGCPTTPKSPKCRAEGRAPRADSGPVSPAASRNGMSVAGNRQSEPRL  872

Qy  670  GLH----HTNPEPHEILLDMGHLGSLALQGEKLHYFDQL--TKGWTGAESDCKQLKKGL  721
Db  873  ASHVAADTAQPRP----TGEKGNIMASDRLERTNQLKIVEISAEAVSETVCGNKPAES  927

Qy  722  DGAAYLLKDGEVK----RLNINQSTSSIKHGTENVFSLLPHVRNKPE-----------  763
Db  928  DRRGGCLAQGNCQEKSEIRLYRQVAESSTSHPS----SLPSHASQAEQEMSRSFSMAKLA  983

Qy  764  -PGDALQ-GLNKDDKAQAMAVI----GV--NKYLALTEKGDIRSFQIKPGTQQLERPAQ  814
           ||                    ||
Db  984  SSSSSLQTAIRKAEYSQGKSSLMSDSRGVPRNSIPGGPSGEDHLYFTPRPATRTYSMPAQ  1043
```

```
Qy  815  TLS---REG----ISGELKDIHV----------------------DHKQ  834
            |||            ||
Db 1044  FSSHFGREGHPPHSLGRSRDSQVPVTSSVVPEAKASRGLPSLANGQGIYSVKPLLDTSR  1103

Qy  835  NLYALTHEGEVF------------HQPREAWQNGAESSSWHKLALPQSESKLk  875
           |     :|||               |     ||::
Db 1104  NLPA-TDEGDIISVQETSCLVTDKIKVTRRHYCYEQNWPH--ESTSFFSV----KQRIK  1155

Qy  876  SLD-MSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGS----------Q  921
           :||                         :|||
Db 1156  SFENLANADRPVA--------KSGASPFLSVSSKPPIGRRSSGSIVSGSLGHPGDAAA  1205

Qy  922  TVFNRLM-----QGVKGVIP--------GSGLTVKLSAQTGGMTGAEGRKVYSSKFSERIRA  970
           |:            :|       :||:
Db 1206  RLLRRSLSSCSENQSEAGTLLPQMAKSPSIMTLTISRQNPPETSSKGS--DSELKKSLGP  1263

Qy  971  YAF-NPTMSTPRPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVRHNAPQPDLQ  1029
                                                        ::::
Db 1264  ----------NKSSVRHTQPSPVSR  1294

Qy 1030  SKLETLDLGEHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSP--  1087
         ||:|                        :|             ||:
Db 1295  SKIQEL----RALSMPDLDKLCSE-DYSA-GPSAVLFKTELEITPRRSPGPPA  1341

Qy 1088  GKALVQSFNVNR------SGQDLSKSLQQAVHATPSAESKLQSMLGHFVSAGVDMSHQKG  1142
         |:                 ||                       :||
Db 1342  GGVSCPEKGGNRACPGGSGPKTSAA-----ETPSSAS  1380

Qy 1143  EIPLGRQRDPNDKTALTKS--------RLILDTV-TIGELHELADKAKLVSDHKPDADQIKQ  1195
           :|                       ::|    |:|||
Db 1381  DLPFRRSWSVNLDQLLVSAGDQQRLQSVLSSVGSKSTILTLQEAKAQSENEEDVCFIVL  1440

Qy 1196  LRQQFDTLREKRYESNPVKHYTDM--GFTHNKALEANYDAVKAFINA-FKKEHHGVN-  1249
                     ::|              |:||
Db 1441  NRKEGSGL----GPSVAGGTDVEPKSITVHRVFSQGAASQEGTMNRGDFLLSVNGASL  1494

Qy 1250  --LTIRTVLESQGSAELAKKLKNTLLSLDSGESMSFSR----------SYGGGVSTVFVPT  1298
            :|                             ||:
Db 1495  AGLAHGNVLKVLHQAQLH--KDALVVIKKG-MDQPRPSARQEPPTANGKGL------  1542

Qy 1299  LSKKVPVPIPGAG-ITLDRAYNLSFSRTSGGLNVSFGRDGGVS  1341
         ||            :|||              ::|||  |||:|
Db 1543  LSRKT-IPLEFGIGRSVAVHDALCVEVLKTSAGLGLSL--DGGKS  1584
```

RESULT 15

US-09-739-449-13178

; Sequence 13178, Application US/09739449
; GENERAL INFORMATION:
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15490)C
; CURRENT APPLICATION NUMBER: US/09/739,449
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 09/514,000
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 13351
; SEQ ID NO 13178
; LENGTH: 1242
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens

US-09-739-449-13178

Query Match   1.7%; Score 163; DB 5; Length 1242;
Best Local Similarity 18.8%; Pred. No. 0.2;
Matches 256; Conservative 146; Mismatches 431; Indels 530; Gaps 62;

```
Qy  282  EGTDTTQSPLKPQSMLKGSGAGVTPLAVTL----DKGKLQLAPDNPPALNTLLKQTLGK  336
          |    ::||   :  :|||   :|      |     :|
Db   28  DGTLVVQGGLQTNSVILGATGGSSGTASFIGANWANDGDVSVGGNGTGSL-LLLHGTTAS  86
```

```
Db   738 ------------------------------------------GTIGGLTVNSGGTLAPGNSIGTLTSTG-NATFASGSTYAVEIDADGS 783
                                                         ||  |||  ||
Qy  1446 SANLDLRAGINLNEDGS---KPNGVTARVSAG--------LSASANL------ 1481
                         |  :|  :||:
Db   784 SDRLAVTGTITIANDVSLIVTPLGAHSAYSLGTRYTILTATGGVTGTFSSVDESFAYLTA 843
             :| :  :  |
Qy  1482 -AAGSRERSTT----------SGQFGSTTSASNNRPTFLNGVGAGANLTAALG------V 1524
             :|     ||   : ||   |||
Db   844 KVAQSGDDATTYLSFLRTSPDSGLLAAATSTANAR-----------AAANAVEALGQTSPLYE 895
           :  :|  |:||
Qy  1525 AHSSTHEGKPVGIFPAFT-STNVSAALALDNRTSQSISLELKR 1566
             :|   | : ::|  |: ::||
Db   896 AALFLQQGETQGAFSQLAGEIHPSLAMALINRSGQSRDVILNR 938
```

Search completed: June 5, 2001, 18:21:11
Job time: 387 sec

```
Qy   337 DTQHYLAHHASSDGSQOHLLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQAGTGSVS 396
           |:     |||  | :|    ||
Db    87 SENIYVGE--SGTGSGSSLKLDNSSTLTVADTVFAGYNSWSSS------------AAGGSGNVE 135
Qy   397 VDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWQHPAGAARPQGESIRLHDDKIHI 456
             |  : | |::|
Db   136 VLG------ASSLYSANGVLANDPDTVGTALVSG------GGSS------ 167
Qy   457 LHPELGVWQSADKDTHSQLSRQADGKIYALKDNRTLQNLSDNKSSEKLVDK--IKSYSVD 514
           |:|  |
Db   168 ------W------------NMSSDLVVGEAGVGTLTIT 187
Qy   515 QRGQVAILTDTPGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISH 574
           |   :  |   |:|
Db   188 GGGMVISQTTT-------VIADLNSADGSSVTVS------RLGSSLQSDSLVVGN 229
Qy   575 G---RLVVADSE----------GKLFSAAIPKQGDGNELKMKAMPQHALDEHFGHDHQISG 622
           |                     :||:|
Db   230 GGAAKLVVEAAATVTSGEAIIGRHSASEATVTGDGSKWTT------GDLQVGG 276
Qy   623 FFHDDHGQLNALVKNNFRQQHACPLGNDHQFHPGWNLTDALVIDNQLGLHHTNEPEHEIL 682
           : ||                      :|:
Db   277 DTSDPGGL-------AGNG-----TLNVTAGGSVDS------ 300
Qy   683 DMGHLGSLALQEGKLHYFDQLTKGWTGAESDCKQLKKGLDGAAYLLKDGE--VKRLNINQ 740
            :  |||:|:                ||   :|     |
Db   301 TVAHLGVVAGATGSA-IVDGKGSVWT---VDRNSLEVGVSGAGSLAVTGGGLVDAANIII 356
Qy   741 STSSIKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGVNKYLALTEKGDIRSF 800
            : | :  |:|
Db   357 GTNTGGNGSVRV-------SGADSTVKSRSDLNVGLY------GN- 388
Qy   801 QIKPGTQQLERPAQTLSREGISGELKDIHVDHKQNLYALTHEGEVFHQPREAWQNGAESS 860
                :  |:|     |||
Db   389 ----GSMTVEAGGAVKSRDG----------YVATYGGST- 413
Qy   861 SWHKLALPQSESKLKSLDMSHEHKPIATFEDGSQHQLKAGGWHAYAAPERGPLAVGTSGS 920
                |                  :  |:    ::| :|
Db   414 ------SAVTVTGDGSSWAMTGTFFVGYAS-----GATGN 442
Qy   921 QTVFNRLMQGVKGKVIPGSGLTVKLSAQTGMTGAEGRKVSSKFSERIRAYAFNPTMSTP 980
             :  |  |  :|  | |:|            :::  ||
Db   443 VTVSN-----GGAIRATGVTL-----GDLAGASGTMITGAGSKVTAYVDNGTVNSG 489
Qy   981 RPIKNAAYATQHGWQGREGLKPLYEMQGALIKQLDAHNVR-HNAPQPDLQSKLETLDLGE 1039
             |                                 |:|
Db   490 ----SVDVGFQGSGSLS---VVNGG--SLDAYNLYVGNA------------LGS 522
Qy  1040 HGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLKSNGEINSEFKPSPGKALVQSFNVNR 1099
            |                     |:|:
Db   523 SGAVL--------VSGVGSHVSV------DGLMVVGNAGN 548
Qy  1100 SGQDLSKSLQQAVHATPPSAESKLQSMLGHFVSAGVDMSHQKGEIPLGRQRDPNDKTALT 1159
            | ||                         :|       |
Db   549 GSVEITGGASLAAPTILIATEAGSTGVLS--IGAG-------SGQTARS 588
Qy  1160 KSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTDM 1219
                 |||:|    ::|                                       :|:
Db   589 AGAVEARAIAFG-----AGNGSIVFNH------SET 613
Qy  1220 GFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAELAKKLKNTLLSLDSGE 1279
            | :|                |  |||     :  || |:
Db   614 GYT----LSADISGAGRVV----AEAGVTTLSGNNSYSGGTTISAGMLKGTAKSFGSGG 664
Qy  1280 SMSFSRSY--GGGVSTVFVPTLSKKVPVPVIPGAGITLDRAYNLSFSRTSGGLNVSFGRD 1337
            ::|  :|   |||                           :    |:|: |
Db   665 IVNNAELVVDGGG--------TLSN------AISGTG------SPEKTG- 693
Qy  1338 GGVSGNIMVATGHDVMPYMTGKKTSAGNASDWLSAKHKISPDLRI-GAAVSGTLQGTLQ 1395
            ||:   |:  ||             |  |             |:  || |
Db   694 --DGNLLL-TGNST--YSGATAVSAGK----LSVNGSLASAVSVGSGATVGGT------ 737
Qy  1396 NSLKFKLTEDELPGFIHGLT----HGTLTPAE---LLQKGIEHQMKQGSKLTFSVD--T 1445
```